

**Caracterización molecular del melanoma
mediante inteligencia artificial:
Implicaciones en pacientes tratados con
inmunoterapia**

TESIS DOCTORAL

Guillermo Prado Vázquez

Programa de Doctorado en Biociencias Moleculares
Universidad Autónoma de Madrid
Madrid, 2020

Departamento de Bioquímica
Facultad de Medicina
Universidad Autónoma de Madrid

Caracterización molecular del melanoma mediante inteligencia artificial: Implicaciones en pacientes tratados con inmunoterapia

Memoria presentada para optar al grado de Doctor por el grado en
Biología

GUILLERMO PRADO VÁZQUEZ

Dirigida por:

Dr. Enrique Espinosa Arranz

Dr. Angelo Gámez Pozo

*Realizada en el Instituto de Investigación Sanitaria del Hospital
Universitario La Paz-IdiPAZ y Biomedica Molecular Medicine S.L.*

*“Entraría en tu luz
Con una canción sencilla
Tres notas y una bandera
Tan blanca como el corazón
Que late en tu cuerpo de niña”*

A Cualquier otra parte

Dorian

Agradecimientos

Al Dr. Espinosa, por siempre arrojar luz con sus preguntas y ser una fuente de conocimiento.

A Juanan, por abrirme las puertas de su laboratorio y luchar día a día por mejorar la calidad de la ciencia.

A Lucía, por su crítica constructiva y su aportación imprescindible en el desarrollo de este trabajo.

A Andrea, por su tesón, lucha y conocimiento del sistema inmune.

A Elena, por siempre ayudarme con su trabajo inestimable.

A Rocío, por su infinita eficacia técnica y porque nunca para de cuidarnos

A todos mis compañeros, os diría que gracias, pero que leches, SALDA DAGO!

A ti Angelo, que decir, muchas gracias por ayudarme día a día, por ser un ejemplo y un referente, por ser más que un jefe un amigo, por luchar para sacarnos a todos adelante, por nuestros bailes y conciertos, por las infinitas horas de discusiones filosófico-científicas, por aguantarme, POR TODO, gracias.

A todas mis amigas, las de toda la vida, las espartans, las biólogas, las pengüins, las croquetts y las subnodramers, por estar siempre a mi lado, por el orgullo que me trasmitís y por hacer mi vida una serie de anécdotas impagable, en especial a mi Bro, por ayudarme a salir vivo de muchas de ellas.

A mi familia por estar ahí, por no dejarme desfallecer, por acogerme de vuelta, por todas esas cosas que se dan por sentado. Papa, Mama, Tete, os quiero.

A mis sobrinos, Martina, Jorge, espero que algún día os leáis esta tesis de vuestro tío y penséis, menuda chorrada, si eso del cáncer ya no preocupa a nadie.

Y por último a ti que por la razón que sea te estás leyendo esto gracias por dedicarme este tiempo.

ABSTRACT

Abstract

Melanoma is the most lethal malignancy of the skin. The molecular analysis of melanoma has improved our understanding of the disease. The Cancer Genome Atlas (TCGA) Network proposed the transcriptomic classification of melanoma in three subtypes: Keratin, MITF-low and Immune. However, this classification has not translated into therapeutic advances yet. The aim of this study is to characterize molecular differences of melanoma at biological and molecular levels. A novel approach based on the existence of different molecular informative layers was used in this study to establish independently sets of information. Probabilistic graphical models followed by successive sparse k-means and consensus cluster analyses was used to classify melanoma tumor samples from the TCGA cohort. Two independent classifications of molecular information were found: an immune and a molecular layer. The molecular classification provided four groups: epidermis development, membrane, melanogenesis low and keratin. The immune classification subdivided the TCGA cohort into an immune-high and an immune-low group. We also studied another cohort of patients treated with immunotherapy mimicking the methods used in the TCGA cohort. The main result was an informative layer that divide the patients in two groups according to their immune response whereas the immune low group had worse prognosis than the immune high. Summarizing, in this work we propose a novel analytical approach based on informative molecular layers. In this way, two independent classifications, an immune-based and a molecular-based classification, were established for melanoma. Information from immune layer in a cohort treated with immunotherapy showed a resistant group with low expression of the immune response genes.

ÍNDICE

Agradecimientos.....	1
Abstract	3
Clave de abreviaturas.....	10
Introducción.....	12
1. Epidemiología	12
2. Factores de riesgo	12
3. Mecanismos y fisiopatología	12
4. Clasificación del melanoma	14
4.1. Clasificación histopatológica	14
4.2. Clasificación genética	15
4.3. Cribado y diagnóstico clínico.....	16
5. Manejo de la enfermedad	16
5.1. Tumor primario.....	16
5.1.1. Terapia adyuvante.....	16
5.2. Enfermedad metastásica	17
5.2.1. Terapia dirigida.....	17
5.2.2. Inmunoterapia	18
6. Clasificaciones moleculares y Atlas del genoma del cáncer (TCGA)	18
7. Metodologías de análisis de datos	20
7.1. Modelos gráficos probabilísticos	20
7.2. K-means y algoritmo Consensus Clúster.....	21
Objetivos.....	23
Material y Métodos	25
1. Cohortes de Pacientes estudiadas	25
1.1. TCGA. Cohorte de descubrimiento.....	25
1.1.1. Procesamiento y filtrado de datos.....	25
2. GEM. Cohorte de validación experimental	26
2.1. Criterios de inclusión.....	26

2.2.	Extracción de ARN	26
2.3.	Análisis de expresión génica por ARN-seq.....	27
2.4.	Preprocesamiento de los datos.....	28
3.	Análisis de datos	29
3.1.	Modelos gráficos probabilísticos (MGP)	29
3.2.	Procesamiento de la red	29
3.3.	Ontología génica.....	30
3.4.	Análisis de nodos.....	30
3.5.	Sparse k-means y análisis de grupos consenso (Capas de información) ..	31
4.	Metodología de identificación de subtipos moleculares	31
5.	Análisis estadísticos.....	31
Resultados		34
1.	Cohorte del TCGA.....	34
1.1.	Base de datos	34
1.1.1.	Información clínica	34
1.1.2.	Base de datos de mARN-seq	35
1.2.	Modelo gráfico probabilístico.....	35
1.3.	Capas de información	36
1.3.1.	Capa 1	37
1.3.2.	Capa 2.....	38
1.3.3.	Capa 3.....	39
1.3.4.	Capa 4.....	40
1.3.5.	Capa 5.....	41
1.3.6.	Capa 6.....	42
1.3.7.	Capa 7.....	43
1.3.8.	Análisis de supervivencia de las capas de información.	43
1.4.	Caracterización de los resultados proporcionados por las capas de información	44
1.4.1.	Clasificación basada en capas inmunes	45

1.4.2.	Clasificación basada en información molecular del tumor.....	47
1.5.	Resumen	50
2.	Cohorte de validación experimental del grupo español multidisciplinar del melanoma	51
2.1.	Clínicos descriptivos	51
2.2.	Análisis de expresión génica por RNA-seq	52
2.3.	Resultados RNA-seq	53
2.4.	Modelo gráfico probabilístico	54
2.5.	Capas de información de la base de datos de descubrimiento.....	54
2.6.	Supervivencia de los grupos de las capas	55
2.6.1.	Capa 2. Biosíntesis de melanina.....	56
2.6.2.	Capa 3. Respuesta inmune	57
2.6.3.	Capa 5. Desarrollo de la epidermis	58
2.6.4.	Capa 7. Homeostasis del ion Calcio	59
2.6.5.	Resumen de las capas.....	60
	Discusión.....	62
1.	Hallazgos de la tesis.....	62
2.	Antecedentes del trabajo	62
3.	Puesta en valor de la metodología de trabajo cohorte publica descubrimiento-cohorte experimental de validación	62
4.	Puntualizaciones técnicas.....	63
5.	Caracterización del MGP	63
5.1	Similitudes y diferencias con MGP de trabajos previos	64
5.2	Procesos interesantes detectados en los nodos	64
5.2.1	Inmunes	64
5.2.2	NADPH oxidasa	65
5.2.3	Desarrollo de la epidermis y queratinización	65
5.2.4	Melanogénesis y melanosoma.....	66
5.2.5	Otros nodos del MGP	67

5.2.6	Resumen del MGP	68
6.	Las capas de información	68
6.1	Consideraciones al análisis de capas	68
6.2	Reorganización de las capas en grupos	69
6.2.1	Racional de este estudio	69
6.2.1.1	Clasificación Inmune.....	69
6.2.1.2	Clasificación Molecular	69
6.3	Comparación con la clasificación transcriptómica del TCGA	70
6.4	Conclusiones del análisis de capas	70
7.	Estudio de validación. Cohorte del GEM	71
7.1	MGP de la base de datos del GEM.....	71
7.2	Capas de información en la base de datos del GEM.....	72
8.	Limitaciones del estudio.....	72
9.	Resumen	74
	Conclusiones	76
	Bibliografía.....	78

CLAVE DE ABREVIATURAS

Clave de abreviaturas

AJCC: (del inglés, American Joint Committee on Cancer): Comité Conjunto Estadounidense sobre el Cáncer.

BIC: Criterio de información bayesiana.

BRAFv600: Gen BRAF mutado en la valina 600.

CAMs (del inglés cell adhesion molecule): Moléculas de adhesión celular.

CCA: (del inglés consensus cluster análisis): Análisis de grupos consenso

FDA (del inglés US Food and Drug Administration): Agencia estadounidense de alimentación y fármacos.

GEM: Grupo español multidisciplinar del melanoma.

GO: Ontología génica.

KEGG: Enciclopedia Kioto de genes y genomas.

M: Presencia de metástasis.

MEC: Matriz extracelular.

MGP: Modelo Gráfico Probabilístico.

N: Afectación ganglionar.

PD1 (del inglés programmed cell death protein 1): Proteína de muerte celular programada 1.

ROS: Especies reactivas de oxígeno.

RPKM: Lecturas de fragmentos por millón de kilobases.

RPPA (del inglés reverse phase protein análisis): Arrays de proteína de fase reversa.

SG: Supervivencia global.

SLP: Supervivencia libre de progresión.

SNP (del inglés single nucleotide polymorfism): Mutaciones puntuales.

T: Tamaño tumoral.

TCGA (Del inglés The Cancer genome Atlas): Atlas del genoma del cáncer.

TME (del inglés, tumor microenvironment): Microambiente tumoral.

UV: Radiación ultravioleta.

WES (del inglés whole exome sequencing): Secuenciación de exoma completo.

WT (del inglés wild type): Fenotipo salvaje.

INTRODUCCIÓN

Introducción

1. Epidemiología

En todo el mundo, se diagnostican unos 232.100 casos de melanoma primario al año. 55.500 de las muertes por cáncer anuales se deben al melanoma cutáneo (0,7%). La distribución de los casos varía a lo largo del mundo en un rango que va desde 0,2 por cada 100.000 personas al año en el sudeste asiático hasta los 35,8 casos por cada 100.000 personas al año en Nueva Zelanda y los 34,9 casos por cada 100.000 personas al año en Australia. En Europa la incidencia es de 10,2 por cada 100.000 personas al año (Torre *et al.*, 2015), cifra que se ha incrementado desde la década de los 70. En cuanto a la mortalidad, va desde 0,1 por cada 100.000 personas al año en el sudeste asiático hasta los 4,7 por cada 100.000 personas al año en Nueva Zelanda o los 4,0 por cada 100.000 personas al año en Australia. La supervivencia a 5 años en Europa desde el 2000 al 2007 va desde 74,3% (Este de Europa) al 87,7% (Norte de Europa). La supervivencia relativa disminuye con la edad de los pacientes, además de ser más alta en mujeres que en hombres. Para melanomas sin invasión de los ganglios linfáticos la supervivencia relativa a cinco años es del 98% en estadios I y 90% en estadios II (Miller *et al.*, 2016).

2. Factores de riesgo

Los factores de riesgo del melanoma cutáneo establecidos son: la radiación ultravioleta producida por la exposición al sol y las quemaduras solares, la presencia de nevus melanocíticos o displásicos, historia personal de melanoma cutáneo, historia familiar de melanoma cutáneo, características fenotípicas como el pelo, los ojos y la piel claros y tendencia de tener pecas y por último un alto estatus socioeconómico (Schadendorf *et al.*, 2018).

3. Mecanismos y fisiopatología

El análisis del mapa genético de distintos tipos de cáncer revela que los melanomas cutáneos son los tumores con mayor carga mutacional (>10 mutaciones por megabase) (figura 1) (Alexandrov *et al.*, 2013).

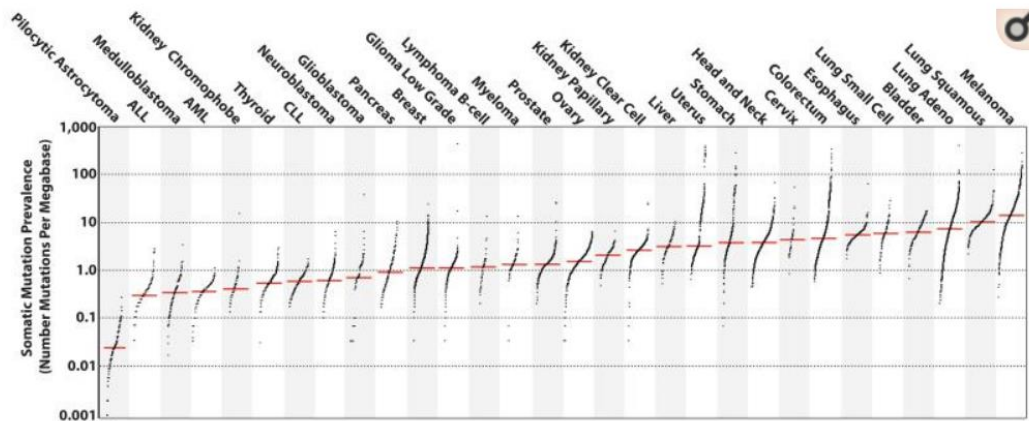


Figura 1. La prevalencia de mutaciones somáticas en los tipos de cáncer humano. Cada punto representa una muestra, mientras que las líneas horizontales rojas son los números medios de mutaciones en los respectivos tipos de cáncer. El eje vertical muestra el número de mutaciones por megabase, mientras que los diferentes tipos de cáncer se ordenan en el eje horizontal en función de su número medio de mutaciones somáticas (Alexandrov *et al.*, 2013).

Muchas de estas mutaciones puntuales se deben a la exposición a la radiación ultravioleta (UV), ya sea directamente por el efecto mutagénico de la radiación UV A o UV B (Hodis *et al.*, 2012; The Cancer Genome Atlas Network, 2015), o de manera indirecta por la producción de radicales libres como resultado de la interacción bioquímica entre UV A y la melanina (Noonan *et al.*, 2012).

El mecanismo principal de transformación oncogénica en melanoma es el resultado de una cascada de señalización celular que termina con la activación constitutiva de un oncogén, como en muchos otros tipos de cáncer. Estudios previos indican que el gen *BRAF* se encuentra mutado en la valina 600 (*BRAF*^{V600}) frecuentemente en los nevos (86.4%) (Kiuru *et al.*, 2018). Se necesita mutaciones adicionales para que estos nevos progresen a melanomas, como pueden ser mutaciones en *TERT*. Además, para ganar potencial invasivo se necesitan mutaciones terciarias que afecten a los genes que controlan el ciclo celular (*CDKN2A*) o la remodelación de la cromatina (*ARID1A*, *ARID1B*, *ARID2*). Por último, parece ser que la progresión metastásica del melanoma está asociada con mutaciones en *PTEN* o *TP53* (Schadendorf *et al.*, 2018).

A nivel proteico, estas alteraciones génicas promueven una sobreestimulación recíproca de las vías celulares afectadas, sobre todo la vías de las *MAPK* y *PI3K*, *AKT*, *PTEN* y *mTOR* (figura 2) (Schadendorf *et al.*, 2018).

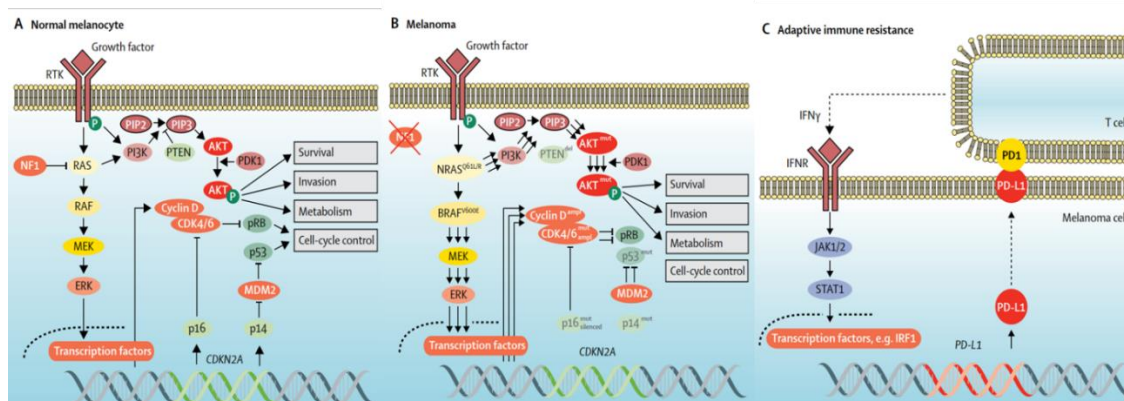


Figura 2. (Schadendorf *et al.*, 2018). Selección de vías de señalización clave y dianas terapéuticas en melanoma. (A) La señalización *MAPK*, *PI3K-AKT* y la regulación del ciclo celular en condiciones normales permiten un control equilibrado de las funciones básicas de la célula. (B) En los melanomas, las alteraciones genéticas conducen a la activación de la vía constitutiva, con pérdida de la homeostasis celular. (C) El punto de control inmunitario *PD-1-PD-L1* está regulado principalmente por la señalización de interferón- γ .

Otro de los procesos claves del cáncer es la evasión del sistema inmune. La teoría de la vigilancia inmunológica propone que los tejidos y las células del cuerpo están constantemente monitorizadas por el sistema inmune y que por tanto este elimina la mayoría de las células cancerígenas incipientes y los tumores surgentes (Ribatti, 2017). Por tanto, para que los tumores crezcan, uno de los primeros pasos que deben dar es evadir el sistema inmune. En 2011 este concepto de evasión del sistema inmune se introdujo entre los marcadores del cáncer clásicos (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011).

Las células de melanoma consiguen evadir el sistema inmune a través de diversos mecanismos. Uno de los mecanismos más importantes es el que implica al *PD1* (programmed cell death protein 1). Tras el reconocimiento del antígeno tumoral por las células T, los interferones liberados desencadenan la expresión de los ligandos de *PD-1* (*PD-L1* y *PD-L2*), mediada por *JAK-STAT*, en la superficie de las células de melanoma. La unión de *PD-L1* y *PD-L2* a *PD-1* conduce a la supresión de la actividad efectora de las células T e inhibe la respuesta inmune antitumoral (resistencia inmune adaptativa) (García-Díaz *et al.*, 2017).

4. Clasificación del melanoma

4.1. Clasificación histopatológica

El diagnóstico del melanoma se realiza de manera histopatológica. Esta clasificación tiene en cuenta el tamaño tumoral (T), la profundidad (índice de Breslow), la afectación ganglionar (N) y la presencia o ausencia de metástasis (Gershenwald *et al.*, 2017). Los datos de T, N y M permiten la estadificación del tumor, desde el estadio I hasta el IV. El estadio IV se refiere a los tumores M1, es decir, los que tienen metástasis (tabla 1).

Cuando T es..	y N es...	y M es..	Entonces el estadio es
Tis	N0 ^b	M0	0
T1a	N0	M0	IA
T1b	N0	M0	IA
T2a	N0	M0	IB
T2b	N0	M0	IIA
T3a	N0	M0	IIA
T3b	N0	M0	IIB
T4a	N0	M0	IIB
T4b	N0	M0	IIC
T0	N1b, N1c	M0	IIIB
T0	N2b, N2c, N3b o N3c	M0	IIIC
T1a/b–T2a	N1a o N2a	M0	IIIA
T1a/b–T2a	N1b/c o N2b	M0	IIIB
T2b/T3a	N1a–N2b	M0	IIIB
T1a–T3a	N2c o N3a/b/c	M0	IIIC
T3b/T4a	Cualquier N \geq N1	M0	IIIC
T4b	N1a–N2c	M0	IIIC
T4b	N3a/b/c	M0	IIID
Cualquier T, Tis	Cualquier N	M1	IV

Tabla 1. Estadios patológicos según el comité americano del cáncer (American Joint Committee on Cancer [AJCC]) en su octava edición (2017).

4.2. Clasificación genética

El consorcio *The Cancer Genome Atlas* (TCGA), a través del análisis de multitud de muestras de melanoma con diversas técnicas moleculares de nueva generación, definió cuatro subtipos de melanoma en función de las mutaciones presentes en cada tumor (The Cancer Genome Atlas Network, 2015). Estos cuatro subtipos son: BRAF mutado, que representa el 50% de los melanomas, N-Ras, K-Ras o H-Ras mutado, que representa aproximadamente el 25% de los melanomas, NF1 mutado, que representa aproximadamente el 15%, y triple *wild-type*, un 10%.

4.3. Cribado y diagnóstico clínico

Las personas con alto riesgo de padecer un melanoma deberían de ser revisadas de manera regular. Normalmente, el cribado del melanoma se basa en una revisión completa del cuerpo por un médico experto, con ayuda de un dermatoscopio. La principal ventaja del cribado por parte de médicos expertos de las lesiones de la piel es que lleva a la detección temprana de los melanomas con poca profundidad (Epstein *et al.*, 1999). Hoy en día hay mucha controversia sobre si el cribado regular de la población es o no necesario. Por un lado, se postula que el cribado regular puede llevar al sobrediagnóstico de melanomas muy finos que no afectan a la supervivencia del paciente (Stang, Jöckel and Heidinger, 2018; Welch, Woloshin and Schwartz, 2005). Por otro, no se ha podido demostrar todavía que los programas nacionales de cribado disminuyan la mortalidad asociada a los cánceres de piel (Stang *et al.*, 2016). Sin embargo, muchas guías clínicas recomiendan que las personas con alto riesgo de melanoma reciban vigilancia regular para mejorar la supervivencia a través de la detección temprana y reducir escisiones innecesarias.

5. Manejo de la enfermedad

5.1. Tumor primario

El tratamiento de la enfermedad localizada consiste en la escisión local con distintos márgenes de seguridad dependiendo del grado de grosor de Breslow (The Cancer Genome Atlas Network, 2015). La linfadenectomía no ha demostrado beneficio en la supervivencia, así que no está recomendada de forma general. La biopsia del ganglio centinela está recomendado para melanomas primarios con un grosor de al menos un milímetro. En el futuro los resultados de los tratamientos neoadyuvantes podrían abrir la posibilidad de una cirugía menos extensa en pacientes que tengan respuestas patológicas completas (Badenas *et al.*, 2012).

5.1.1. Terapia adyuvante

La recurrencia del melanoma después de la cirugía está muy relacionada con el estadio patológico al diagnóstico, y es mayor en el estadio III, es decir, en aquellos tumores con ganglios regionales afectos. En estos tumores, el tratamiento adyuvante debe ser considerado para prevenir la recaída y mejorar la supervivencia global (Schadendorf *et al.*, 2018).

El estándar actual de tratamiento adyuvante para el estadio III se basa en tres ensayos clínicos. Uno de ellos demostró la ventaja de la combinación dabrafenib y trametinib frente a placebo en el caso de tumores con mutación en *BRAF* (Eggermont *et al.*, 2016).

Los otros dos estudios demostraron la utilidad de los anti-PD1 nivolumab (frente a ipilimumab) y pembrolizumab (frente a placebo), independientemente del estado mutacional de *BRAF* (Yokota *et al.*, 2019; Weber *et al.*, 2017).

5.2. Enfermedad metastásica

El melanoma en estadio IV es aquel que presenta metástasis a distancia del tumor primario al momento del diagnóstico. El melanoma puede invadir prácticamente cualquier órgano, si bien los sitios de metástasis más frecuentes son los ganglios linfáticos, los pulmones, los huesos y el cerebro (Damsky, Rosenbaum and Bosenberg, 2010). Hay dos tipos de tratamientos principales para estos tumores, terapia dirigida e inmunoterapia.

5.2.1. Terapia dirigida

Para melanomas localmente avanzados y metastásicos con mutaciones en *BRAF* están aprobados en primera línea los inhibidores orales de *BRAF* y *MEK*. Algunos ensayos clínicos han demostrado una respuesta objetiva a inhibidores de *BRAF* de aproximadamente el 50%, pudiendo aumentar esta respuesta hasta un 70% cuando se combina con inhibidores de *MEK*. La mediana de supervivencia libre de progresión aumenta con la terapia combinada de *BRAF* y *MEK* hasta los 11-14.9 meses, por lo que la combinación se ha convertido en el tratamiento dirigido estándar (Schadendorf *et al.*, 2018). Estos tratamientos producen una rápida respuesta en los pacientes, de entre unos días a unas pocas semanas, independientemente de la carga tumoral o la localización de las metástasis. La mediana de supervivencia global de los pacientes con melanoma en estadio IV tratados con inhibidores de *BRAF* y *MEK* está entre los 22-25 meses, con un 40% de los pacientes alcanzando los 3-5 años (Ascierto *et al.*, 2016).

El principal problema de la terapia dirigida es la aparición de resistencia secundaria. Se han identificado múltiples mecanismos de resistencia a los inhibidores de *BRAF* y *MEK* que llevan a la activación de la ruta de las *MAPK* o de la ruta *PI3K-AKT*. Entre estos mecanismos, la amplificación del gen de *BRAF* y mutaciones en *MEK1* y *MEK2* son las mejor descritas (Van Allen *et al.*, 2014; Long *et al.*, 2014).

En cuanto a los pacientes con melanoma sin mutaciones en *BRAF*, las opciones de terapias dirigidas son escasas. Existen algunos ensayos clínicos con pacientes *NRAS* mutados tratados con inhibidores de *MEK* (binimetinib) que muestran un 15% de tasa de respuesta total y una mediana de supervivencia libre de progresión 2,8 meses. Aunque este resultado es superior a lo que conseguiría una quimioterapia clásica, sigue siendo modesto, por lo que la terapia dirigida no se considera una buena opción para pacientes con tumores no mutados en *BRAF*.

5.2.2. Inmunoterapia

Uno de los principales avances que se ha producido en el tratamiento del melanoma avanzado es la inmunoterapia frente a los puntos de control inmune *CTLA4* (ipilimumab) y *PD1* (pembrolizumab y nivolumab). Los primeros estudios en el campo demostraron que el tratamiento con ipilimumab incrementaba el porcentaje de largos supervivientes (más allá de 3 años) en pacientes con melanomas metastásicos en torno a un 20% (Schadendorf *et al.*, 2015). Por otro lado, nivolumab y pembrolizumab demostraron una eficacia mayor cuando se compararon con quimioterapia y con ipilimumab, respectivamente y (Larkin *et al.*, 2015). Estos dos fármacos anti-Pd1 fueron aprobados para su uso en melanoma metastásico, y las guías clínicas los recomiendan antes que el ipilimumab (Michielin *et al.*, 2019). También se ha aprobado la combinación de nivolumab más ipilimumab, que produce una tasa de respuestas superior a la de los anti-PD1 en monoterapia pero a costa de una elevada toxicidad.

Algo más de la tercera parte de los pacientes muestran resistencias *de novo* a la terapia anti PD1, y un 20-30% de los respondedores iniciales adquieren también resistencia. Una vez más, el conocimiento de estos mecanismos de resistencia resulta crucial para poder tratar mejor a estos pacientes. El desarrollo de biomarcadores que ayuden a predecir de forma temprana la respuesta a estos tratamientos inmunoterápicos es fundamental, tanto para evitar una toxicidad innecesaria en los pacientes que no se vayan a beneficiar de la misma, como para explorar alternativas terapéuticas. Algunos biomarcadores propuestos son la carga tumor o la expresión de neoantígenos (Van Allen *et al.*, 2015), la presencia de un microambiente tumoral inmuno-permisivo con expresión de *PD-L1*, presencia de linfocitos infiltrados en el tumor o actividad de la ruta de señalización del interferón- γ (Schadendorf *et al.*, 2018). Sin embargo, hoy en día no existe ningún biomarcador suficientemente validado para su uso en rutina clínica (Michielin *et al.*, 2019).

6. Clasificaciones moleculares y Atlas del genoma del cáncer (TCGA)

La mayoría de los estudios moleculares en grandes series de tumores busca mecanismos de resistencia a fármacos (Riaz *et al.*, 2017), grupos moleculares (Lehmann *et al.*, 2011), o perfiles con valor pronóstico y predictivo (Pu *et al.*, 2017). Diversos grupos de investigación han hecho un esfuerzo enorme en este campo, aunque pocas veces se alcanzan resultados que permitan tomar decisiones en la clínica (Buono *et al.*, 2018). Uno de los grupos que más investigaciones con impacto clínico está logrando es el consorcio The Cancer Genome Atlas (TCGA). Hasta la fecha han analizado cohortes de más de 30 tipos de tumores (Tomczak, Czerwińska and Wiznerowicz, 2015). Los estudios de cada tipo de

cáncer, así como los análisis que estudian varios tipos juntos, han ampliado el conocimiento actual de la tumorigénesis. Un objetivo principal del consorcio es hacer públicas las bases de datos generadas, para ayudar a mejorar los métodos de diagnóstico, los estándares de tratamiento y, finalmente, para prevenir el cáncer. Aplicando una metodología de análisis similar en la mayoría de sus trabajos, entre las que se encuentran la secuenciación de exoma completo (WES), los *arrays* de mutaciones puntuales (SNP), el estudio del perfil de metilación del ADN y *arrays* de proteína de fase reversa (RPPA), han establecido las bases de los subtipos moleculares en diversos tipos de cáncer, como el cáncer de mama (The Cancer Genome Atlas Network, 2012), cáncer de vejiga (Rodríguez-Vida, Lerner and Bellmunt, 2018), cáncer renal (Linehan and Ricketts, 2019) o el adenocarcinoma de pulmón (The Cancer Genome Atlas Network, 2018) entre otros.

En 2015 el TCGA publicó su investigación en melanoma (The Cancer Genome Atlas Network, 2015). En este artículo se describe el mapa de las alteraciones genómicas que sufren los melanomas cutáneos a través del estudio del ADN, ARN y las proteínas de 333 melanomas primarios y/o melanomas metastásicos de 331 pacientes. A través de estos análisis establecieron la clasificación molecular de la que ya hemos hablado; BRAF mutado, RAS mutado, NF1 mutado y triple WT (fenotipo salvaje). Estos cuatro grupos no tenían diferencias en su supervivencia global, calculada desde el día de colección de la muestra hasta el día de la muerte o la fecha de último seguimiento. No obstante, en este trabajo se definieron dos grupos de pacientes según la infiltración de linfocitos medida de manera semicuantitativa y la expresión de la proteína LCK, que sí guardaban relación con la supervivencia de los pacientes (figura 3).

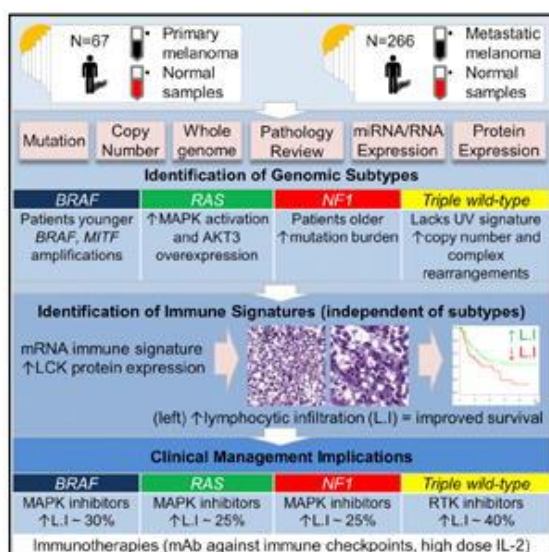


Figura 3. Esquema resumen de los resultados presentados en (The Cancer Genome Atlas Network, 2015).

Sin embargo, como reflejan las guías clínicas consultadas, estas clasificaciones no han llegado a impactar en el manejo clínico de la enfermedad (Michielin *et al.*, 2019).

Como ya hemos comentado previamente, la mayoría de los trabajos del TCGA se basan en una metodología de análisis de datos estándar (Levan *et al.*, 2010). En el caso de la clasificación “transcriptómica”, basada en los análisis de secuenciación del ARNm, se genera una clasificación en tres grupos a partir de un análisis de agrupamiento jerárquico (HCL). Un subtipo “inmune”, otro subtipo de “Queratinas” y un último subtipo “MITF-low”. El subtipo inmune tiene una serie de genes sobreexpresados asociados a células inmunes, moléculas de señalización inmunes, proteínas co-estimulantes o co-inhibidoras inmunes, citocinas, quimioquinas y sus correspondientes receptores. El subtipo “Queratinas” se caracteriza por alta expresión de genes relacionados con queratinas, pigmentación y epitelio, así como con genes asociados al desarrollo neuronal y embriológico. Este grupo mostró peor supervivencia. El subtipo “MITF-low” está caracterizado por la expresión baja de algunos genes asociados con pigmentación y epitelio (The Cancer Genome Atlas Network, 2015).

En este trabajo reevaluaremos esta clasificación utilizando una metodología de análisis de datos diferente y novedosa (los modelos gráficos probabilísticos) que ya ha sido desarrollada por nuestro grupo en otras patologías (Trilla-Fuertes *et al.*, 2020b; Trilla-Fuertes *et al.*, 2019a; Trilla-Fuertes *et al.*, 2019b; Prado-Vázquez *et al.*, 2019; de Velasco *et al.*, 2017).

7. Metodologías de análisis de datos

7.1. Modelos gráficos probabilísticos

Los modelos gráficos probabilísticos (MGP) son modelos gráficos indirectos que buscan minimizar el criterio de información bayesiana (BIC), desarrollados por Gideon E. Schwarz (Schwarz, 1978). Para esto se siguen dos pasos: se busca el árbol de máxima verosimilitud y a partir de ahí, en pasos sucesivos se introducen nuevas aristas revisando si eso reduce el BIC de manera que se preserve la capacidad de descomposición del gráfico inicial (Markowetz and Spang, 2007).

Los MGP permiten descripciones o atribución de funciones por inferencia (Markowetz and Spang, 2007; Schäfer and Strimmer, 2005). Este tipo de modelos permiten, a partir de los datos de expresión de genes y sin necesidad de ninguna otra información, construir una red que relacione los genes por sus patrones de expresión (Markowetz and Spang, 2007). Experiencias previas del laboratorio han demostrado la utilidad de esta aproxi-

mación a la hora de analizar datos moleculares con orígenes ómicos distintos a la expresión génica, como proteómica (Trilla-Fuertes *et al.*, 2020a; Rojas L *et al.*, 2019; Gámez-Pozo *et al.*, 2017a; de Velasco *et al.*, 2017), microARNs (Gámez-Pozo *et al.*, 2015; Trilla-Fuertes *et al.*, 2020b; Gámez-Pozo *et al.*, 2017a) o metabolómica (Trilla-Fuertes *et al.*, 2019a), permitiendo evaluar funcionalmente diversos tipos de cáncer, como cáncer de mama (Prado-Vázquez *et al.*, 2019; Trilla-Fuertes *et al.*, 2019a; Trilla-Fuertes *et al.*, 2018; Zapater-Moros *et al.*, 2018; Gámez-Pozo *et al.*, 2017a), de vejiga (Trilla-Fuertes *et al.*, 2019b; de Velasco *et al.*, 2017), de ano (Trilla-Fuertes *et al.*, 2020a), renal (Trilla-Fuertes *et al.*, 2020b) e incluso melanoma (Trilla-Fuertes *et al.*, 2019c).

7.2. K-means y algoritmo Consensus Clúster

El estudio de enfermedades mediante el uso de datos ómicos se ha convertido en un enfoque popular que potencialmente puede conducir a una medicina de precisión (Huo and Tseng, 2017). El análisis *sparse k-means* permite establecer grupos diferenciales entre muestras y obtener información sobre las características moleculares específicas de cada grupo de tumores (Trilla-Fuertes *et al.*, 2018). Por otro lado, el algoritmo Consensus Clúster es capaz de definir el número óptimo de grupos que puede proporcionar una clasificación molecular (de Velasco *et al.*, 2017).

OBJETIVOS

Objetivos

En los últimos años se ha realizado un esfuerzo importante para caracterizar molecularmente el melanoma, pero hasta ahora la mayor parte de ese esfuerzo no ha repercutido en la atención clínica de los pacientes que padecen dicha enfermedad.

Nuestro grupo ha desarrollado una metodología de análisis de datos innovadora, basada en modelos gráficos probabilísticos y análisis *sparse k-means* /análisis de grupos consenso recursivo, que permite una interpretación de la información molecular más orientada hacia la clínica.

Sobre estas premisas, y trabajando con los datos de pacientes de melanoma disponibles, así como con datos de una cohorte de validación, nos planteamos los siguientes objetivos:

1. Caracterización funcional de los datos de melanoma disponibles en la cohorte del TCGA
 - a. Análisis mediante modelos gráficos probabilísticos
 - b. Búsqueda de clasificaciones moleculares definiendo capas de información mediante *sparse k-means* / análisis de grupos consenso recursivo
2. Evaluación de la información obtenida en el primer objetivo en una cohorte independiente, esta vez un grupo de pacientes tratados con inmunoterapia.

MATERIAL Y MÉTODOS

Material y Métodos

1. Cohortes de Pacientes estudiadas

Se utilizaron dos cohortes de pacientes en este estudio. La primera es la cohorte del TCGA y la segunda es la cohorte del grupo español del melanoma (GEM). La cohorte del TCGA se utilizó como cohorte de descubrimiento y se realizó un análisis de datos empleando modelos gráficos probabilísticos (MGP) y análisis de capas de información. Por otro lado, la cohorte del GEM se utilizó como cohorte de validación experimental y se replicaron los análisis tanto el MGP como de capas de información.

1.1. TCGA. Cohorte de descubrimiento

La plataforma <http://firebrowse.org/> contiene diversos datos de pacientes con melanoma recogidos por el consorcio del TCGA. A saber, datos clínicos, variantes de SNPs, mutaciones, metilaciones, expresión de microARNs, expresión de ARNm por secuenciación y *arrays* de proteínas de fase reversa. Se descargaron los datos tanto clínicos como de secuenciación de ARNm de la cohorte.

Los datos clínicos se organizaron por columnas teniendo en cuenta los siguientes parámetros: días hasta el último seguimiento, año de inicio de la patología, supervivencia libre de enfermedad en meses, estatus de la supervivencia libre de enfermedad (si ha recurrido/progresado o sigue sin enfermedad), los meses hasta el éxito y si ha fallecido o no el paciente (supervivencia global), el tipo de muestra (tumor primario o metástasis), el tamaño tumoral (T), la afección de los ganglios (N), la presencia o no de metástasis al diagnóstico (M), si ha recibido tratamiento, tanto neoadyuvante como adyuvante, y el subtipo mutacional (BRAF, NRAS, NF1 o triple negativo), además de los parámetros del paciente relativos al género, edad y raza.

Los datos de secuenciación de ARNm en firebrowse se descargaron parcialmente procesados. Como había muestras analizadas con distintas plataformas, se seleccionaron únicamente las muestras que tenían datos con origen en el equipo HiSeq de Illumina. Al final se obtiene una matriz con las lecturas de fragmentos por millón de kilobases (RPKM) para cada gen, que está anotado tanto con su *gene symbol* como con su valor de Entrez.

1.1.1. Procesamiento y filtrado de datos

A partir de la base de datos de secuenciación de ARNm se aplicaron unos criterios de calidad. Se eliminaron las muestras que presentaban valores de expresión en menos del

50% de los genes. Se filtraron aquellos genes con más de un 25% de valores perdidos. Además, se transformaron los datos usando el logaritmo en base 2 y se imputaron los valores perdidos a la cola de la distribución de los datos, es decir, se generaron valores aleatorios para cada valor perdido. Para este último paso se usó la herramienta Perseus usando los criterios predefinidos del software: los datos perdidos se imputan con datos aleatorios que presentan una distribución normal de anchura $0,3 \cdot \text{desviación estándar}$ y con el centro desplazado $1,8 \cdot \text{desviación estándar}$ hacia abajo respecto al centro de la distribución total de los datos. Estos criterios aseguran que los valores generados para sustituir a los valores perdidos se sitúan en la parte inferior de la distribución de los datos, y de esta forma no generan un sesgo en los análisis posteriores debido a la presencia de valores muy pequeños y alejados de los valores medios de la serie.

Se redujo el número de genes seleccionando aquellos genes con variabilidad mayor en la base de datos. Se consideró como genes variables aquellos que presentaban una varianza mayor de 1,6. La base de datos resultante se empleó para la construcción del modelo gráfico probabilístico y para los análisis posteriores.

2. GEM. Cohorte de validación experimental

2.1. Criterios de inclusión

A través de la colaboración con el Grupo Español Multidisciplinar del Melanoma (GEM) se reclutaron pacientes que cumpliesen los siguientes criterios de inclusión:

1. Pacientes con melanoma avanzado irresecable o metastásico.
2. Pacientes con melanoma cutáneo o de mucosas.
3. Pacientes que hayan recibido tratamiento con nivolumab o pembrolizumab (independientemente de la línea de tratamiento) en el contexto de la práctica asistencial habitual o que lo estén recibiendo (al menos que hayan recibido tratamiento con estos fármacos hasta la progresión de la enfermedad o durante un mínimo de 6 meses si en el momento de recogerse datos aún siguen en tratamiento).
4. Pacientes en los que se disponga biopsia archivada de tejido metastásico.
5. Pacientes vivos o fallecidos de los que se disponga de la información clínica.

2.2. Extracción de ARN

Las muestras tumorales se obtuvieron fijadas en formol e incluidas en parafina. Los bloques de parafina se cortaron con un microtomo Microm HM350S. Se hicieron entre 10 y

15 cortes de 3µm de cada bloque, dependiendo del tamaño de este. El ARN total se aisló con el kit miRNeasy FFPE Kit de Qiagen, siguiendo las instrucciones del fabricante. Brevemente, las muestras se desparafinaron con solución de desparafinado. Después mediante una serie de pasos de incubación y centrifugado se separan restos de tejido insolubles. Posteriormente se incubó con DNasa. El sobrenadante se transfirió a tubos *RNeasy MinElute spin column*, de los que finalmente se eluye el ARN aislado. El control de cuantificación y purificación de ARN extraído se llevó a cabo con un espectrómetro ND-1000 NanoDrop (Thermo Fisher Scientific).

2.3. Análisis de expresión génica por ARN-seq

Para la preparación de las librerías de ARN se empleó el kit *KAPA RNA Hyperprep* (Roche NimbleGen Inc), que es un kit de preparación de librerías mediante cizallamiento. Partiendo de 100 ng de ARN por muestra, primero se hizo un paso de fragmentación por calor en presencia de magnesio. Una vez fragmentado el ARN, se sintetizó la primera hebra de ADNc usando cebadores aleatorios. La síntesis de ADNc de la segunda hebra convierte el híbrido ADNc:ARN a ADNc de doble cadena, mientras va marcando la segunda hebra con dUTP, De manera que se añade el dAMP al extremo 3' de los fragmentos de cADN de doble cadena. Por último, se ligaron los adaptadores 3'dTMP a los 3'dAMP de los fragmentos de la librería. Tras unos pasos de lavado a base de esferas, el ADN de la librería unido al adaptador se amplificó mediante PCR mientras que la hebra marcada con dUTP no se amplificó. En este caso, se aumentó el número de ciclos de PCR recomendados hasta 14 ciclos debido a la baja integridad del ARN las muestras, al provenir de bloques de parafina, para lograr una cantidad suficiente de cADN para la captura. Por último, la distribución de los fragmentos de la librería se confirmó mediante electroforesis. La concentración de la librería se determinó mediante qPCR, usando el kit *KAPA Library Quantification* (Roche NimbleGen Inc) (figura 4).

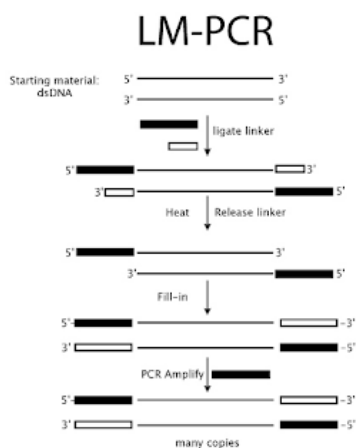


Figura 4. Esquema resumen de los pasos empleados para la generación de librería usando LM-PCR.

El siguiente paso fue la hibridación del ADNc de la librería con el pool de la captura de ARN (SeqCap EZ probe pool, Roche). Para ello, previamente se diseñó esta captura empleando la plataforma informática NimbleDesign de Roche. Se fijó un tamaño máximo de región a capturar de 7 megabases y se seleccionaron genes de interés (ver resultados). Los genes de la captura se preseleccionaron a partir del MGP de la base de datos del TCGA y de una serie de biomarcadores de células del sistema inmune (Angelova *et al.*, 2015). Después se realizó una ligación de los adaptadores de secuenciación masiva con índices específicos para cada muestra. Las librerías así obtenidas se purificaron con AMPure XP beads (KAPA Biosystems), y se amplificaron con cebadores complementarios a la secuencia de los adaptadores. Una vez así generadas, las librerías se mezclaron en cantidades equivalentes y se hibridaron con las sondas de ADN biotiniladas previamente diseñadas, capaces de capturar y enriquecer las regiones deseadas (transcritos de interés). Después de esta hibridación se realizaron una serie de lavados para eliminar los fragmentos de ácidos nucleicos inespecíficos y el producto resultante se volvió a amplificar por PCR. Tras una cuantificación final por fluorimetría y un control de calidad de los tamaños de las librerías, éstas están listas para secuenciar.

Una vez realizadas las capturas de ARN se realizó la secuenciación de las muestras en dos grupos en un equipo NextSeq 500 de Illumina, empleando un cartucho SP con formato de lectura *pair end* 2x100 NextSeq 500/550 High Output Reagent Cartridge v2, 75 ciclos. A continuación el mapeo lo hicimos con el programa TopHat y el test de expresión diferencial basado en el cálculo de FPKM se hizo con el programa Cufflinks (Ghosh and Chan, 2016). Todos esos programas están integrados dentro de la suite G-PRO de BioTechvana (Futami *et al.*, 2011).

2.4. Preprocesamiento de los datos

Se creó una secuencia de comandos en R para sacar los valores de expresión de los genes preseleccionados en la captura de ARN de entre todas las lecturas. Esta secuencia de comandos se basa en los paquetes “reshape2” y “dplyr” de R. La secuencia de comandos importa la matriz “input”, que contiene los datos que se generan tras la secuenciación, de manera que se pueda hacer una búsqueda de los genes preseleccionados en la captura por su *gene symbol* mediante la construcción de un vector de búsqueda. Una vez obtenidos los genes coincidentes se seleccionaron únicamente aquellos genes para los que más de un 50% de los pacientes presentaban expresión por encima del umbral de detección.

Una vez obtenida la base de datos con los genes a estudiar, se realizaron los mismos pasos de preprocesamiento empleados anteriormente en la base de datos del TCGA. Es decir,

se transformó la base de datos haciendo un logaritmo en base 2 y se imputaron los valores perdidos a la cola de la distribución normal empleando también los criterios establecidos en la herramienta Perseus (anchura $0,3 \times$ desviación estándar y con el centro desplazado hacia abajo $1,8 \times$ desviación estándar). La base de datos resultante se utilizó al completo para la construcción de un modelo gráfico probabilístico.

Por último, se utilizaron los genes coincidentes entre las dos bases de datos (TCGA y GEM) para replicar el análisis de las capas de información del TCGA en la base de datos del GEM y permitir su comparación entre ambas bases de datos.

3. Análisis de datos

3.1. Modelos gráficos probabilísticos (MGP)

Para realizar los modelos gráficos probabilísticos se empleó una secuencia de comandos de R (“gRapHD”). La secuencia de comandos recibe como entrada la base de datos y tras convertirla en una matriz busca el árbol de máxima verosimilitud en un primer paso. A partir de ahí en pasos sucesivos introduce nuevas aristas revisando si eso reduce el criterio de información bayesiana (BIC) de manera que se preserve la descomponibilidad del gráfico inicial. Al final el output son dos archivos de texto que resultan en una serie de coordenadas. Esto es un archivo que contiene tres columnas, la primera contiene el número de la interacción, las columnas dos y tres contienen el nombre de dos variables (genes) que están relacionadas. Este archivo de texto se importó en el programa, Cytoscape V3.7.2, que permite su representación gráfica (Shannon *et al.*, 2003).

3.2. Procesamiento de la red

Una vez importado el archivo de texto en el programa Cytoscape, obtuvimos una red compuesta por una serie de nodos (cajas) cada uno de los cuales es representación de un gen. Los nodos están unidos por una línea (arista, que refleja una relación entre ellos, pero no una direccionalidad o dependencia ya que estos modelos gráficos son modelos no dirigidos.

Lo primero que se hizo en los modelos gráficos fue emplear una distribución de Cytoscape (*organic layout*) para separar de manera clara cada uno de los nodos. Una vez separados se comienza a elegir desde fuera a dentro conjuntos de genes por “ramas” para estudiar su ontología génica.

3.3. Ontología génica

La ontología de un gen se refiere a los atributos de un producto génico en cualquier organismo estudiado a tres niveles, la función molecular del gen, la localización de la función del gen en componentes celulares y su rol en procesos biológicos.

Se utilizó la herramienta web DAVID para encontrar las ontologías de las “ramas” de la red (Huang, Sherman and Lempicki, 2009b). Se descartó la categoría CC porque presenta resultados que no aplican en este estudio y se añadieron las categorías de rutas de señalización de las bases de datos de la enciclopedia Kioto de genes y genomas (KEGG) y PANTHER.

Se introdujeron cada uno de los conjuntos de genes (identificados con *gene symbol*) en la herramienta. DAVID devuelve un archivo que contiene las ontologías principales de cada conjunto de genes. Para ello el software realiza un test exacto de Fisher comparando el número de genes del input que tienen una determinada ontología con el número total de genes que están catalogados dentro de esa misma ontología en la base de datos de referencia (en este caso, la base de datos anotada de *Homo sapiens*) (Huang, Sherman and Lempicki, 2009a). La *p* de la prueba se corrige para el multitesteo con las correcciones de Bonferroni y Benjamini, y se proporciona una información del enriquecimiento relativo en veces de cambio. De esta manera se observan las ontologías sobrerrepresentadas en cada “rama”. La ontología con mayor sobrerrepresentación y significancia estadística de cada conjunto de genes se denomina ontología principal de la “rama”.

Una vez analizadas todas las ontologías de las “ramas” de la red se unieron aquellas que tenían la misma ontología principal y que, además, estaban próximas en el MGP y se separaron aquellos conjuntos de genes que tenían ontologías génicas sobrerrepresentadas muy diversas. Este es un proceso de ida y vuelta que se ha de hacer en cada red de manera manual. Una vez terminado se obtuvieron una serie de conjuntos de genes cada uno de ellos con una ontología principal, que se denominan nodos funcionales (Gámez-Pozo *et al.*, 2015).

3.4. Análisis de nodos

Una vez definida la estructura funcional, la actividad de los nodos funcionales se calculó mediante el promedio de la expresión de los genes de cada nodo relacionados con la ontología principal. De esta forma, se obtiene un valor resumen de cada nodo funcional de cada muestra.

Estos nodos funcionales son una medida resumen definida en por el grupo que sirve para dividir las redes en un número reducido de variables que reflejen los grandes procesos biológicos relevantes en la enfermedad (Gámez-Pozo *et al.*, 2017a).

3.5. Sparse k-means y análisis de grupos consenso (Capas de información)

Se utilizó el *Sparse k-means* para definir y caracterizar los grupos de melanoma. El *sparse k-means* asigna un peso a cada variable, según su relevancia en la clasificación de la muestra (Witten and Tibshirani, 2010). En este caso, cada variable es un gen, al cual se le asigna un peso en función de la relevancia que tiene a la hora de dividir la cohorte en grupos de pacientes. Luego, se aplicó un análisis de grupos consenso (*consensus cluster* CCA) utilizando únicamente las variables que fueron seleccionadas por el método de *sparse k-means* para definir el número óptimo de grupos para cada caso (Witten and Tibshirani, 2010). De esta forma se obtiene lo que se denomina una capa de información, que se define como una base de datos que contiene unos genes seleccionados por el *sparse k-means* y una serie de muestras clasificadas en grupos a través del CCA. Una vez identificada una capa de información se eliminan los genes de la base de datos inicial y se realiza de nuevo el análisis para buscar capas subyacentes.

Esta metodología se empleó primero en la base de datos del TCGA para obtener las capas de información. Además, dado que los genes de cada una de las capas de información de la base de datos del TCGA están representados en su mayoría en la base de datos del GEM, se trasladó la información de una base de datos a otra. Para ello se seleccionaron exclusivamente los genes de cada capa y se realizó el CCA para definir la clasificación optima de las muestras.

4. Metodología de identificación de subtipos moleculares

Para comprobar las clasificaciones definidas por las capas o para definir clasificaciones a partir del resultado de varias capas de información se construyeron agrupamientos jerárquicos utilizando la distancia euclídea y el método de *average-linkage* para calcular la distancia entre grupos. El *average-linkage* entre dos grupos se calcula como la distancia media entre los componentes de cada grupo (Nielsen, 2016).

5. Análisis estadísticos

Los análisis estadísticos se realizaron utilizando GraphPad v8. El análisis del MGP se realizó en el programa Cytoscape 3.7.2. Los agrupamientos jerárquicos se definieron utilizando la distancia euclídea en la herramienta MeV (Nielsen, 2016). Las curvas de supervivencia se estimaron mediante el análisis de Kaplan-Meier y se compararon con la

prueba de rango logarítmico (log-rank), utilizando la supervivencia global (SG) y la supervivencia libre de progresión (SLP) como punto final. La SLP se definió como el tiempo entre el día de inicio de tratamiento y la fecha de progresión radiológica o la última fecha de seguimiento. La SG se definió como el tiempo desde el inicio de tratamiento hasta la fecha de exitus o la última fecha de seguimiento. Las correlaciones se evaluaron mediante la r de Pearson y regresión lineal. Las diferencias en la actividad de los nodos funcionales entre los grupos se evaluaron mediante la prueba de Kruskal-Wallis y las comparaciones múltiples se evaluaron mediante la prueba de comparaciones múltiples de Dunn. Los diagramas de caja y bigotes son diagramas de caja de Tukey. Todos los valores de p son bilaterales y $p < 0,05$ se considera estadísticamente significativo. Los datos de expresión y los análisis de red se realizaron en paquetes de software MeV y Cytoscape.

RESULTADOS

Resultados

1. Cohorte del TCGA

1.1. Base de datos

La información de expresión génica de las muestras de melanoma analizadas por el TCGA, así como la información clínica de los pacientes de los que provienen estas muestras se obtuvieron del repositorio firebrowse.org.

1.1.1. Información clínica

La base de datos clínica incluye 470 pacientes diagnosticados de melanoma. Se descartaron tres pacientes, uno de ellos porque no presentaba datos de ARN-seq, y dos porque, aunque estaban presentes, no tenía valores de expresión en más del 50% de los genes.

Tras este filtro, la base de datos clínica que se ha empleado en los análisis posteriores incluye 467 pacientes con las siguientes características clínicas (tabla 2).

Características de los pacientes		Número	Porcentaje
Género	Mujeres	180	39%
	Hombres	287	61%
Tratamiento neoadyuvante	No disponible	432	93%
	Quimioterapia	2	0,4%
	Vacuna/inmunoterapia	3	1%
	Interferón	29	6%
	Otros	1	0,2%
Melanoma primario conocido al diagnóstico	No	47	10%
	Sí	420	90%
Múltiples melanomas primarios al diagnóstico	No disponible	73	16%
	No	382	82%
	Sí	12	3%
Edad (intervalos)	No disponible	8	2%
	<20	6	1%
	21-30	20	4%
	31-40	40	9%
	41-50	76	16%
	51-60	108	23%
	61-70	92	20%
	71-80	86	18%
	>80	31	7%
AJCC Tamaño tumoral (T)	No disponible	29	6%
	T0	23	5%
	T1	10	2%
	T1a	21	4%
	T1b	10	2%
	T2	32	7%
	T2a	30	6%
	T2b	15	3%
	T3	14	3%
	T3a	39	8%
	T3b	37	8%

	T4	15	3%
	T4a	25	5%
	T4b	112	24%
	Tis	8	2%
	T desconocido	47	10%
AJCC Infiltración linfática (N)	No disponible	21	4%
	No	235	50%
	N1	17	4%
	N1a	18	4%
	N1b	38	8%
	N2	6	1%
	N2a	13	3%
	N2b	21	4%
	N2c	9	2%
	N3	54	12%
	N desconocido	35	7%
AJCC Metástasis (M)	No disponible	28	6%
	Mo	415	89%
	M1	5	1%
	M1a	4	1%
	M1b	5	1%
	M1c	10	2%
Tratamiento Adyuvante	No disponible	49	10%
	No	326	70%
	Si	92	20%
Subtipo Mutante	No disponible	106	23%
	BRAF	189	40%
	NF1	24	5%
	NRAS	91	19%
	Triple Negativo	57	12%

Tabla 2. Características de los pacientes de la base de datos del TCGA. La columna número hace referencia al número de pacientes con cada característica y la columna porcentaje hace referencia al porcentaje de pacientes de cada característica en relación con el total de pacientes de la base de datos del TCGA.

1.1.2. Base de datos de mRNA-seq

Se descargó la base de datos de mRNA del repositorio firebrowse.org. La base de datos elegida fue `lluminahiseq_rnaseqv2-RSEM_genes_normalized` y contenía expresión de 20.531 genes para 469 muestras de melanoma. En un primer paso de filtrado se eliminaron 29 genes que no estaban identificados con *gene symbol*. Posteriormente se evaluó el porcentaje de genes con valor de expresión para cada muestra. Después se filtraron aquellos genes con más de un 25% de valores perdidos en toda la serie, pasando la base de datos a tener 15.963 genes. El último paso de filtrado fue transformar los datos usando el logaritmo en base 2 y se imputaron los valores perdidos.

1.2. Modelo gráfico probabilístico

Para construir un modelo gráfico probabilístico (MGP) interpretable es necesario reducir el número de variables, en este caso genes, de la base de datos. Se redujo el número de genes tomando como criterio de variabilidad una varianza mayor de 1.6, que pasaron 2.971 genes. Se construyó un MGP a partir de esta base de datos. Una vez realizado el

MGP se analizó la ontología génica de cada una de sus ramas mediante la herramienta web DAVID (Huang, Sherman and Lempicki, 2009b), dando como resultado el modelo de la figura 5, que consta de 21 nodos funcionales. Para cada paciente se calcula la actividad de cada nodo funcional, como una medida resumen de la expresión promedio de los genes que tienen una misma ontología génica.

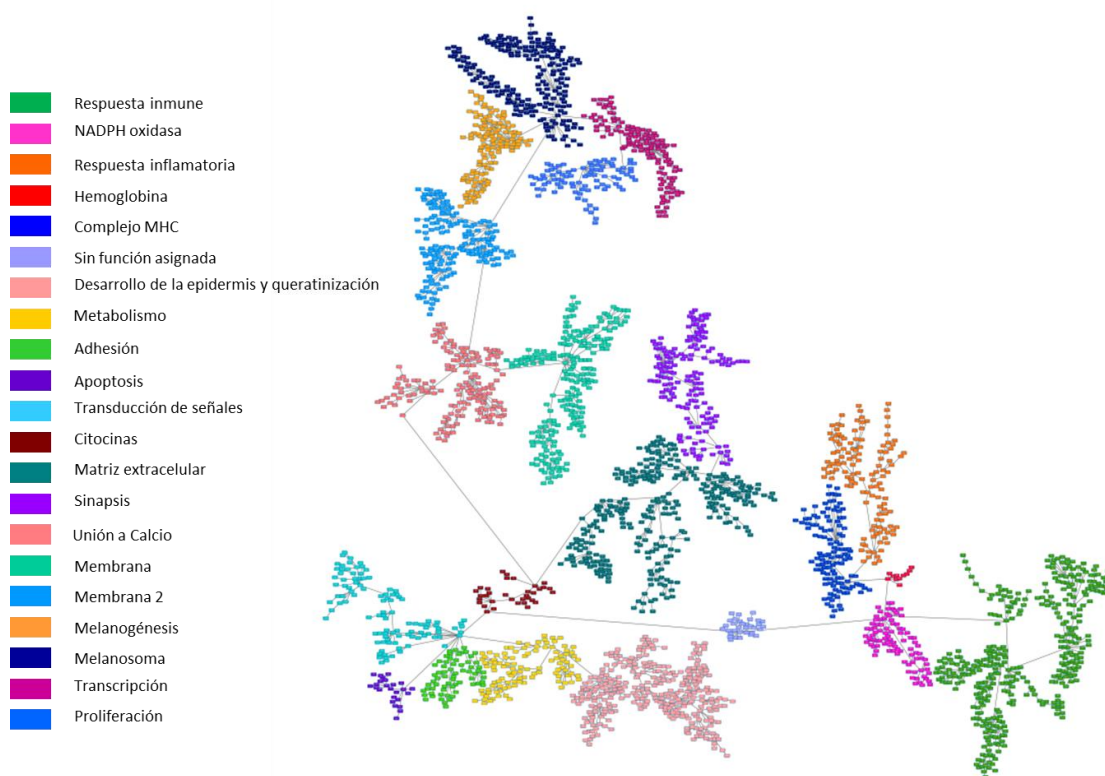


Figura 5. MGP a partir de los 2.971 genes más variables de la base de datos del TCGA. Cada nodo funcional tiene un color diferente comprobable en la leyenda de la parte izquierda de la figura.

1.3. Capas de información

A través de la aplicación recursiva de los algoritmos *sparse k-means* y análisis de grupos consenso (CCA) se obtuvieron siete capas de información (tabla 3).

Capa de información	Genes	Principal Ontología génica	Número óptimo de grupos
1	119	Respuesta inflamatoria	2
2	56	Melanosoma	3
3	150	Respuesta inmune	2
4	106	Ruta de señalización PI3K-Akt	2
5	48	Desarrollo de la epidermis	4
6	129	Respuesta inmune	2
7	70	Espacio extracelular	2

Tabla 3. Capas de información obtenidas de la base de datos del TCGA. Se presenta la información sobre el número de genes, la ontología génica principal y el número óptimo de grupos.

1.3.1. Capa 1

La primera capa de información cuya ontología principal es respuesta inflamatoria está compuesta por 119 genes esparcidos por todo el MGP (figura 6).

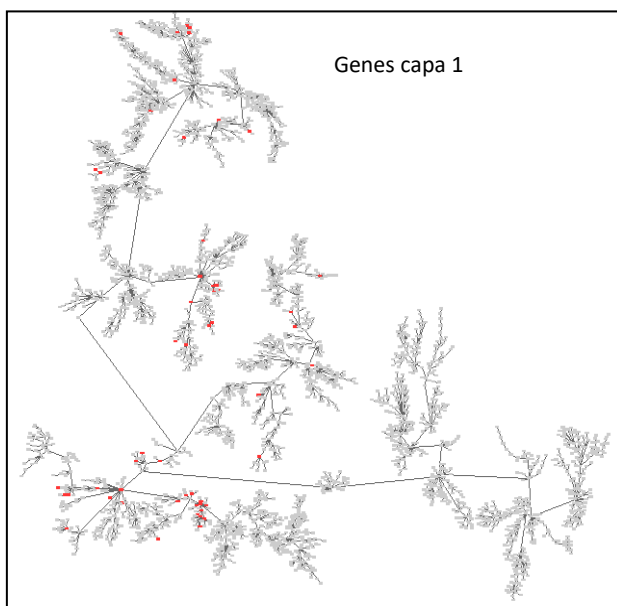


Figura 6. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. En rojo se representan los genes de la primera capa de información.

Los pacientes se dividieron en dos grupos según el CCA. La actividad de los nodos del MGP revela diferencias significativas ($p < 0.05$) entre los dos grupos en los nodos que tienen relación con la respuesta inmune, a saber: inmune, respuesta inflamatoria, citocinas y complejo MHC, y en los nodos de adhesión, metabolismo, NADPH y hemoglobina, (figura 7).

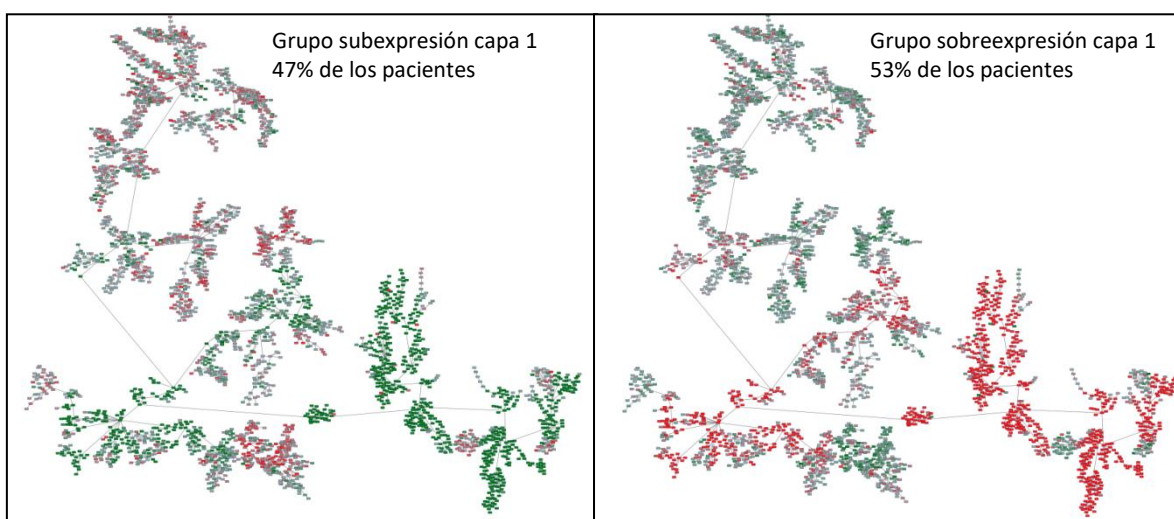


Figura 7. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. La expresión de los genes centrada a la media de expresión se representa en un mapa de calor que va de verde (-1) a rojo (+1) pasando por gris (0).

1.3.2. Capa 2

La segunda capa de información cuya ontología principal fue melanosoma está compuesta por 56 genes, los cuales se sitúan en su mayoría en los nodos de melanogénesis y melanosoma del MGP (figura 8).

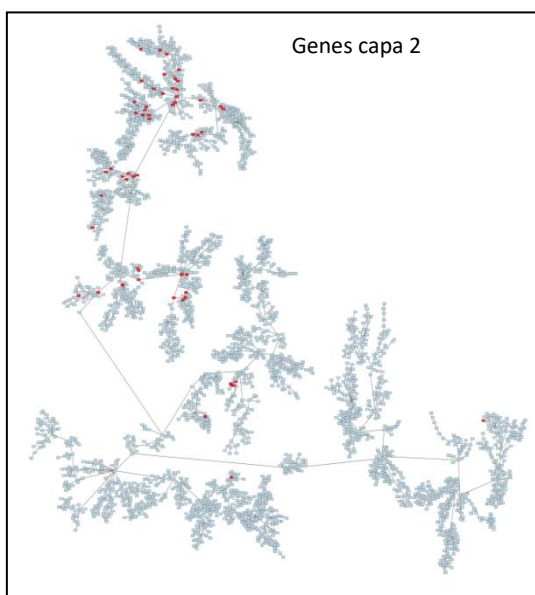


Figura 8. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. En rojo se representan los genes de la segunda capa de información.

Los pacientes se dividieron en 3 grupos según el CCA con diferente expresión de los nodos del MGP (figura 9). El grupo 1 tiene mayor actividad del nodo de proliferación que los otros dos grupos ($p < 0.0001$ frente a ambos), el grupo 2 tiene mayor actividad del nodo de melanosoma que los otros dos grupos ($p < 0.001$ frente al grupo 1 y $p < 0.0001$ frente al grupo 3) y el grupo 3 tiene mayor actividad del nodo de transcripción que los otros dos grupos ($p < 0.0001$ frente a ambos).

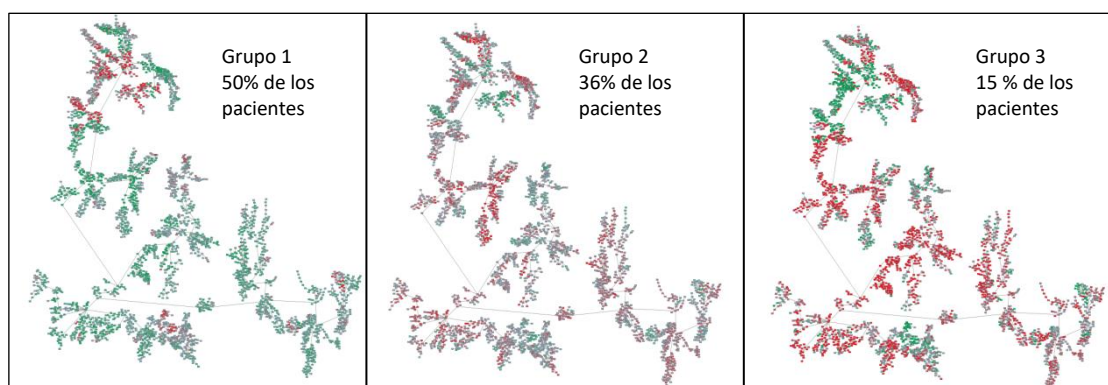


Figura 9. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. La expresión de los genes centrada a la media de expresión se representa en un mapa de calor que va de verde (-1) a rojo (+1) pasando por gris (0).

1.3.3. Capa 3

La tercera capa de información cuya ontología principal es respuesta inmune está compuesta por 150 genes, en su mayoría situados en los nodos de respuesta inmune, respuesta inflamatoria y complejo MHC (figura 10).

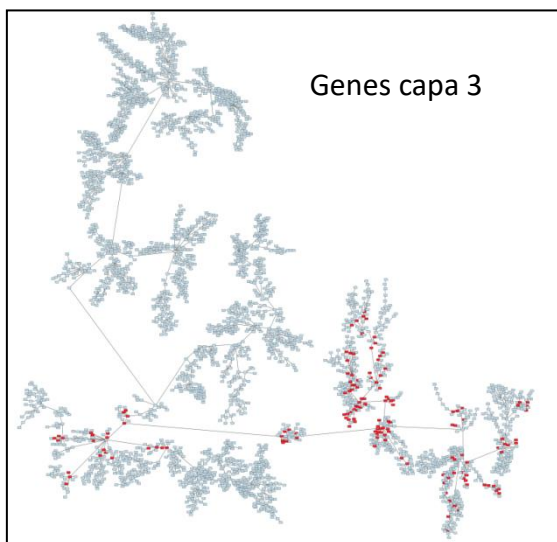


Figura 10. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. En rojo se representan los genes de la tercera capa de información.

Los pacientes se dividieron en 2 grupos según el CCA con diferente expresión de los nodos del MGP (figura 11). El grupo 2 tiene sobreexpresión de los genes de la capa y mayor actividad de los nodos de respuesta inmune, respuesta inflamatoria, complejo MHC y membrana extracelular ($p < 0.0001$).

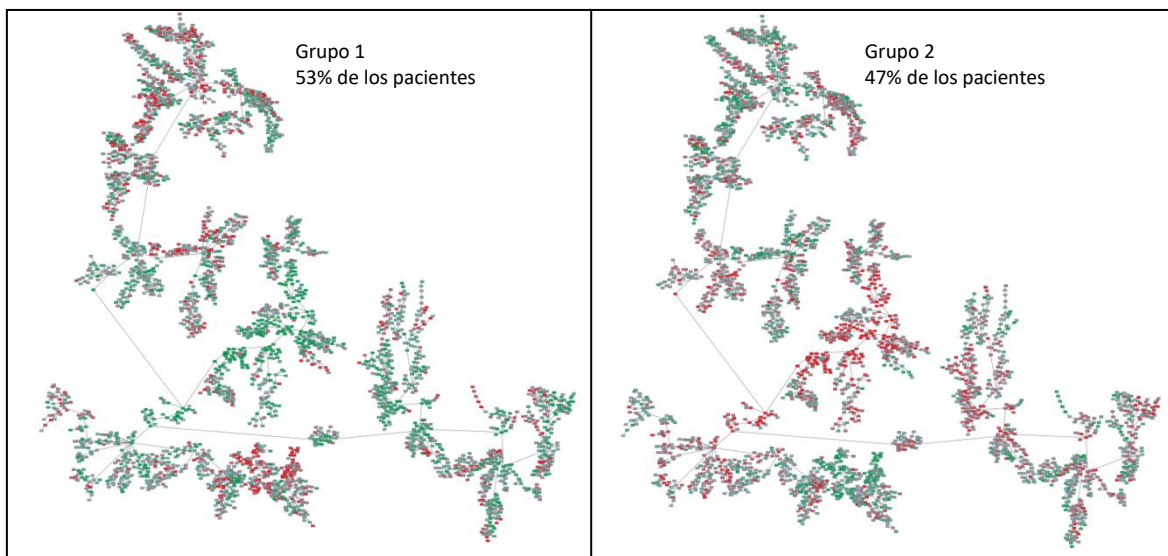


Figura 11. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. La expresión de los genes centrada a la media de expresión se representa en un mapa de calor que va de verde (-1) a rojo (+1) pasando por gris (0).

1.3.4. Capa 4

La cuarta capa de información cuya ontología principal es la ruta de señalización *PI3K-Akt* está compuesta por 106 genes esparcidos por todo el MGP (figura 12).

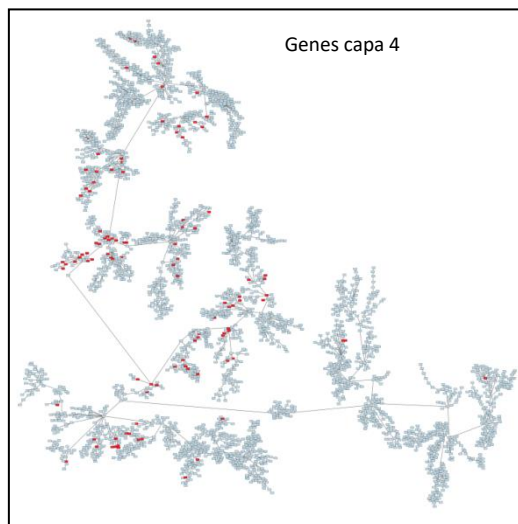


Figura 12. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. En rojo se representa los genes de la cuarta capa de información.

Los pacientes se dividieron en dos grupos según el CCA con diferente expresión de los nodos del MGP (figura 13). El grupo 1 tiene mayor actividad de los nodos de proliferación y melanogénesis que el grupo 2 ($p < 0.001$).

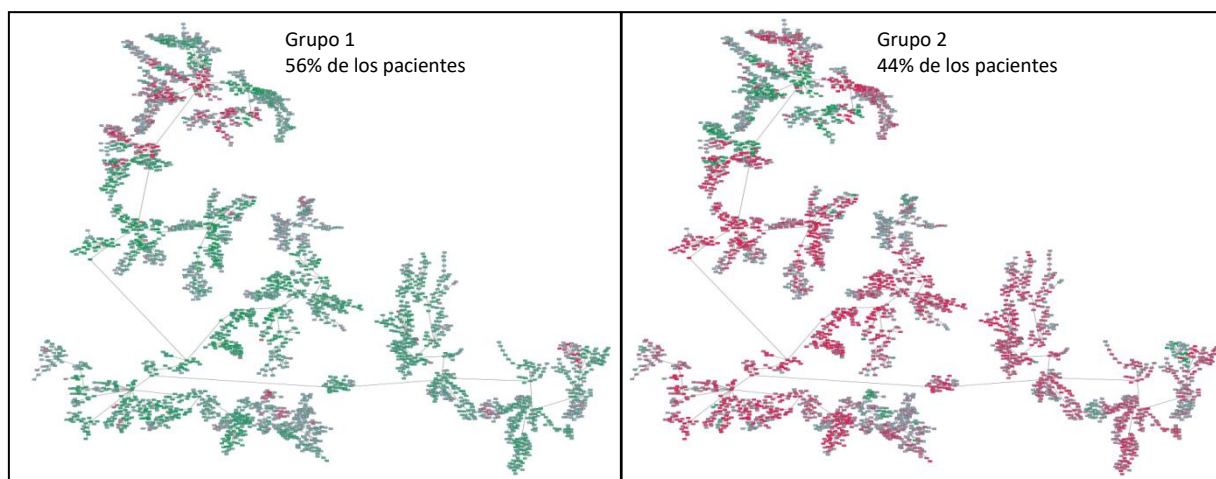


Figura 13. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. La expresión de los genes centrada a la media de expresión se representa en un mapa de calor que va de verde (-1) a rojo (+1) pasando por gris (0).

1.3.5. Capa 5

La quinta capa de información, cuya ontología principal es desarrollo de la epidermis incluye 48 genes situados en el nodo de desarrollo de epidermis y queratinas (figura 14).

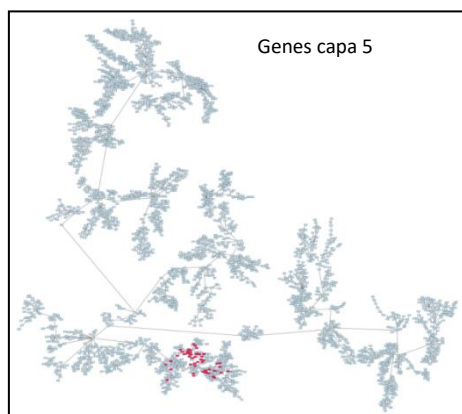


Figura 14. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. En rojo se representa los genes de la quinta capa de información.

Los pacientes se dividieron en cuatro grupos según el CCA con diferente expresión de los nodos del MGP (figura 15). El grupo 4 tenía la mayor actividad de los nodos inmunes (respuesta inmune, MHC, hemoglobina, citocinas y respuesta inflamatoria), además de los nodos de adhesión y apoptosis ($p < 0.0001$). El grupo 3 tenía la mayor actividad del nodo de desarrollo de epidermis y queratinas ($p < 0.0001$).

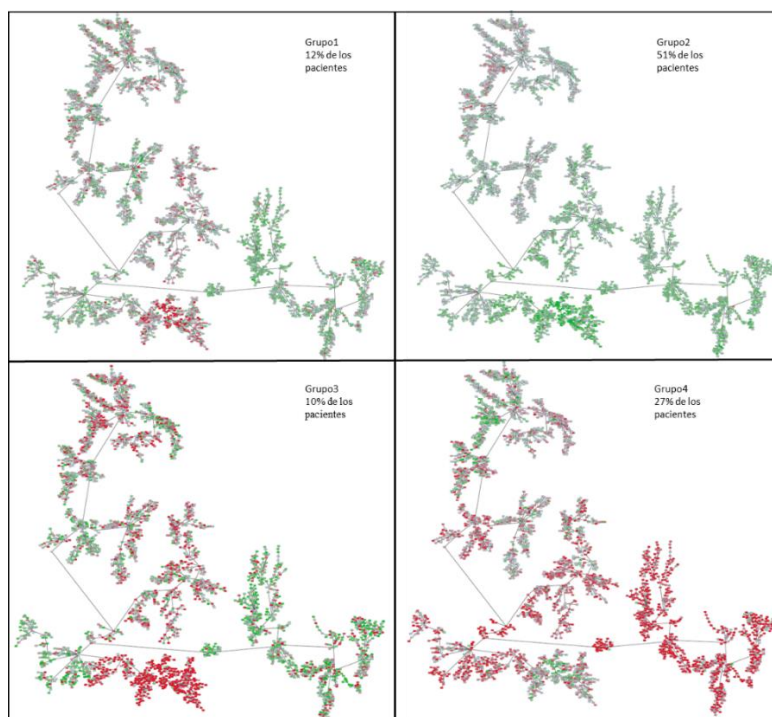


Figura 15. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. La expresión de los genes centrada a la media de expresión se representa en un mapa de calor que va de verde (-1) a rojo (+1) pasando por gris (0).

1.3.6. Capa 6

La sexta capa de información cuya ontología principal respuesta inmune es está compuesta por 129 genes situados en el nodo los nodos de respuesta inmune, NADPH oxidasa, respuesta inflamatoria, complejo MHC, citocinas y transducción de señales (figura 16).

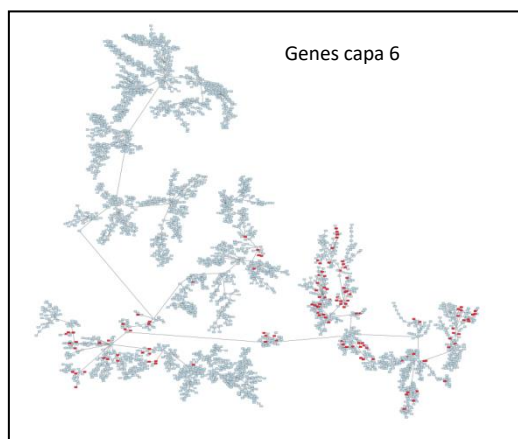


Figura 16. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. En rojo se representa los genes de la sexta capa de información.

Los pacientes se dividieron en 2 grupos según el CCA con diferente expresión de los nodos del MGP (figura 17). El grupo 2 tiene mayor actividad de los nodos de respuesta inmune, respuesta inflamatoria y complejo MHC que el grupo 1 ($p < 0.0001$).

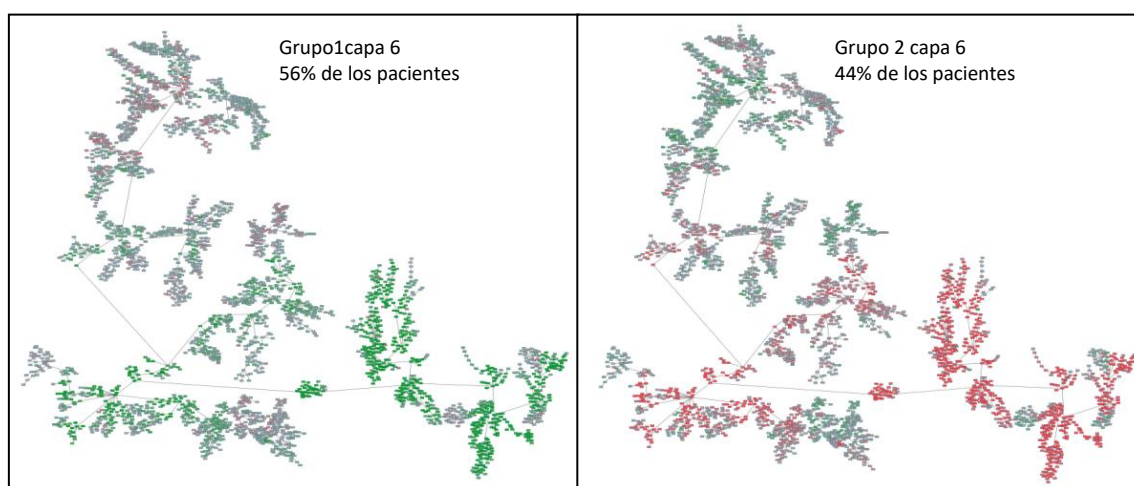


Figura 17. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. La expresión de los genes centrada a la media de expresión se representa en un mapa de calor que va de verde (-1) a rojo (+1) pasando por gris (0).

1.3.7. Capa 7

La séptima y última capa de información cuya ontología principal es espacio extracelular está compuesta por 70 genes, repartidos por el MGP (figura 18).

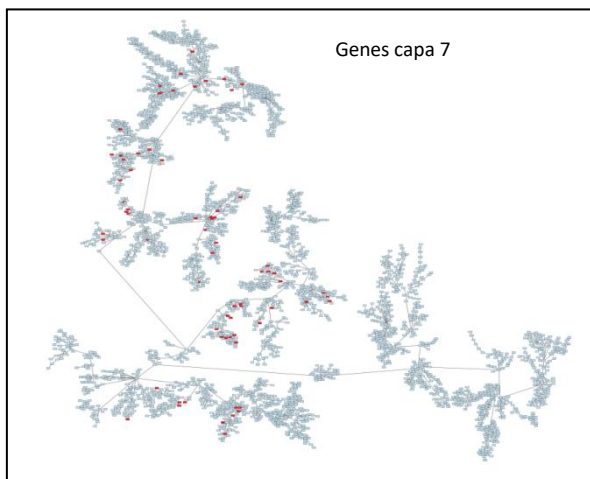


Figura 18. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. En rojo se representa los genes de la séptima capa de información.

Los pacientes se dividieron en 2 grupos según el CCA con diferente expresión de los nodos del MGP (figura 19). El grupo 2 presenta una mayor actividad de los nodos de desarrollo de epidermis y membrana que el nodo 1 ($p < 0.001$).

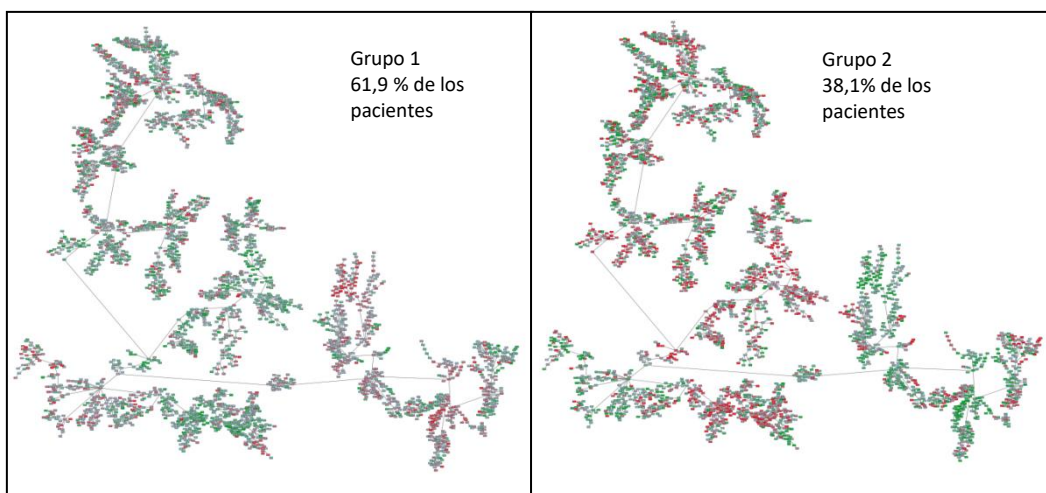


Figura 19. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. La expresión de los genes centrada a la media de expresión se representa en un mapa de calor que va de verde (-1) a rojo (+1) pasando por gris (0).

1.3.8. Análisis de supervivencia de las capas de información.

Solo se apreciaron diferencias significativas en la supervivencia libre de enfermedad entre los grupos identificados en la segunda capa, donde el grupo 3 tiene mejor supervivencia libre de enfermedad que los otros grupos con una mediana de supervivencia libre de

enfermedad de 80 meses frente a los 50 meses del grupo 1 y los 44,58 meses del grupo 2 (figura 20).

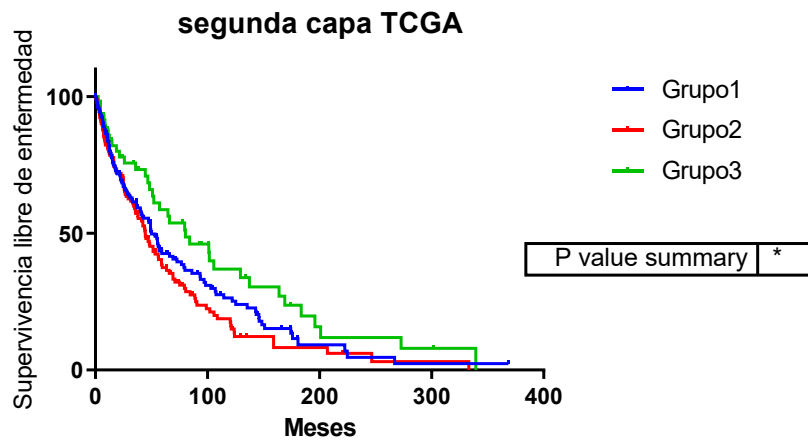


Figura 20. Curvas de supervivencia libre de enfermedad de los tres grupos de la segunda capa de información.

1.4. Caracterización de los resultados proporcionados por las capas de información

Una vez descritas las capas, se construyó un agrupamiento jerárquico con la base de datos del TCGA completa para ver las semejanzas entre los grupos de pacientes en cada capa (Figura 21).

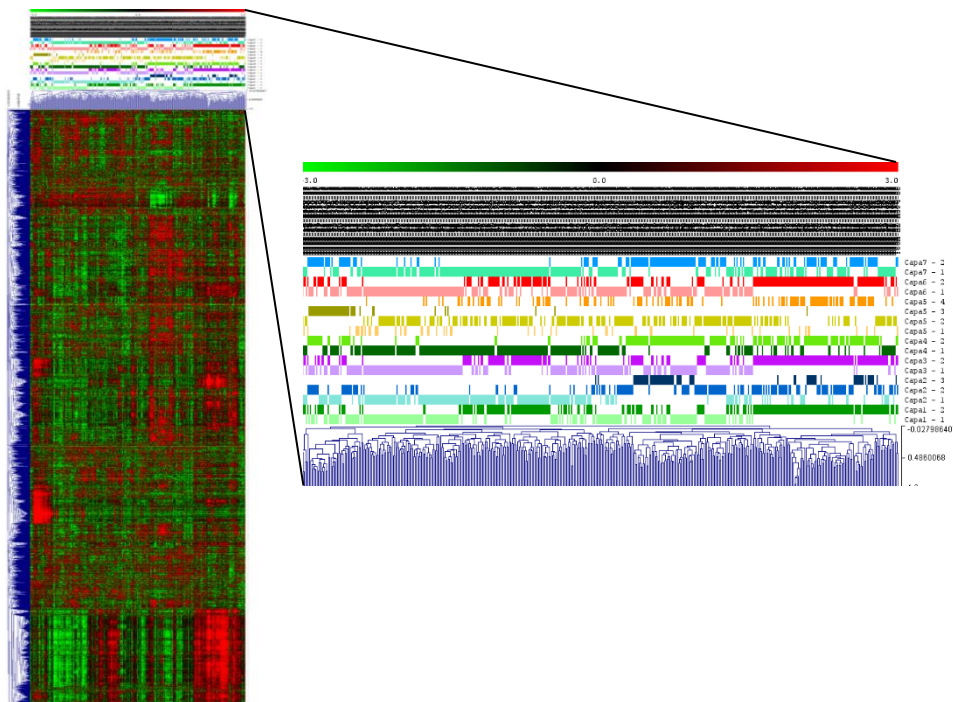


Figura 21. Agrupamiento jerárquico de la base de datos del TCGA representa a los 472 pacientes y las filas a los 2.971 genes (filas). En colores están representados los grupos de las siete capas de información.

El agrupamiento jerárquico demostró que las clasificaciones de las capas inmunes (1,3 y 6) dividían a los pacientes en dos grupos muy similares, los pacientes del grupo con subexpresión de los genes de cada una de las tres capas coinciden en un 94% y los pacientes del grupo de sobreexpresión de los genes de cada una de las tres capas coinciden en un 79%. Asimismo, las demás capas (2, 4, 5 y 7) reflejan una posible división de los pacientes en cuatro grupos en el agrupamiento jerárquico (figura 22).

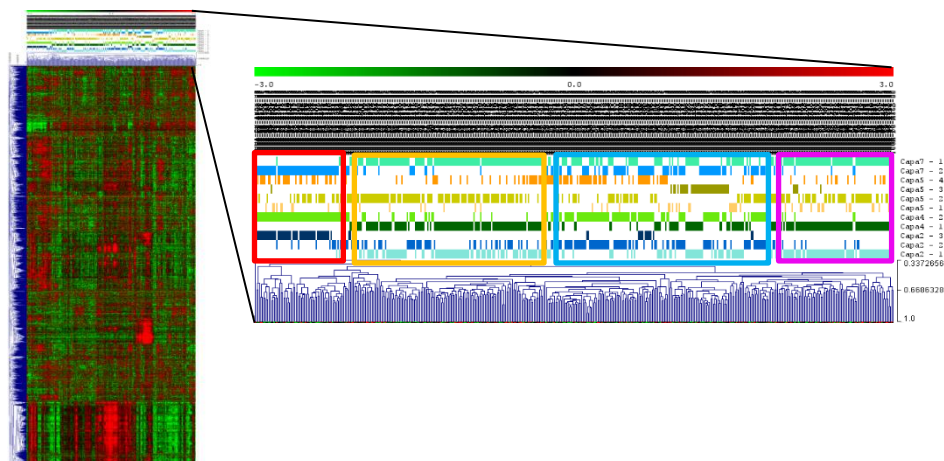


Figura 22. Agrupamiento jerárquico de la base de datos del TCGA. Se representa un agrupamiento jerárquico para los 472 pacientes (columnas) y otro para los 2.971 genes (filas). En colores están representados los grupos de las siete capas de información 2, 4, 5 y 7. Se recuadran los posibles cuatro grupos de pacientes que surgirían a partir de estas cuatro capas.

A partir de las capas de información se realizaron dos clasificaciones una basada en las capas inmunes y otra basada en las demás capas que tenían que ver con el contexto molecular del tumor.

1.4.1. Clasificación basada en capas inmunes

Las capas de información 1,3 y 6 mostraron en el agrupamiento jerárquico una posible clasificación en dos grupos. Utilizando los 398 genes de estas tres capas (figura 23) se realizó un CCA que dividió a los pacientes en 2 grupos.

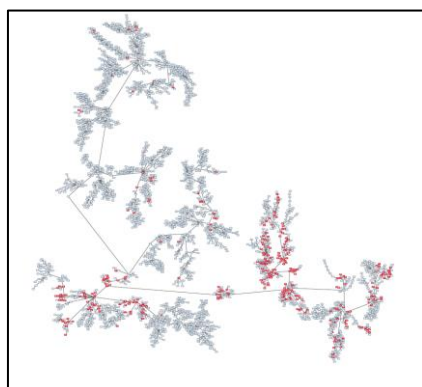


Figura 23. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. En rojo se representa los genes de las capas de información 1,3 y 6.

El grupo 1 está compuesto por 232 pacientes (49%), mientras que 238 pacientes pertenecen al grupo 2 (51%). Estos 2 grupos se representaron en un HCL con los 398 genes de las capas (figura 24). Se puede observar en el HCL que el grupo 1 tiene sobreexpresión de los genes inmunes y el grupo 2 sub-expresión de estos.

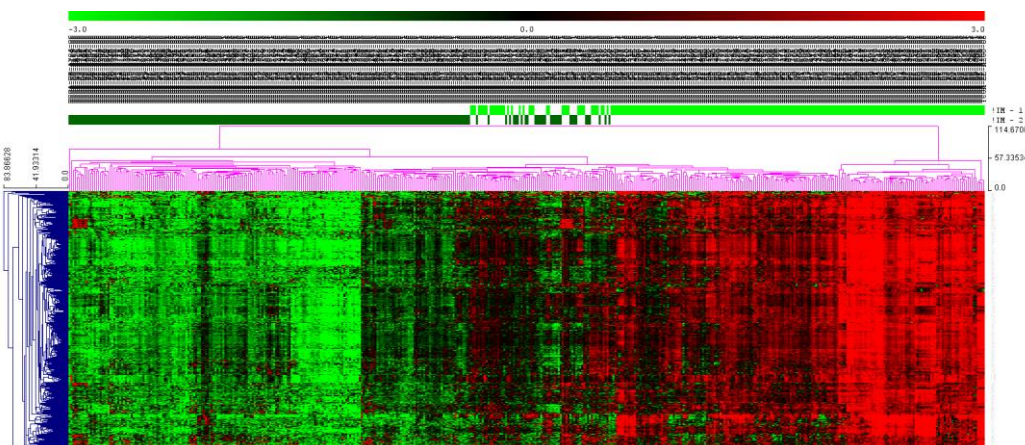


Figura 24. Agrupamiento jerárquico de la base de datos del TCGA. Se representa un HCL para los 472 pacientes (columnas) y otro para los 398 genes (filas). En colores están representados los grupos de la clasificación inmune.

Para poder realizar una comparación entre ambos grupos se calcularon las actividades de los veintiún nodos del MGP, así como un mapa de calor del MGP (Figuras 25 y 26). El primer grupo tiene sobreexpresión de los nodos de respuesta inmune, respuesta inflamatoria y complejo MHC. El grupo 2 sin embargo tiene baja expresión de estos nodos. Se realizó el análisis de supervivencia y se advirtió que no hay diferencias en supervivencia libre de enfermedad entre ambos grupos inmunes.

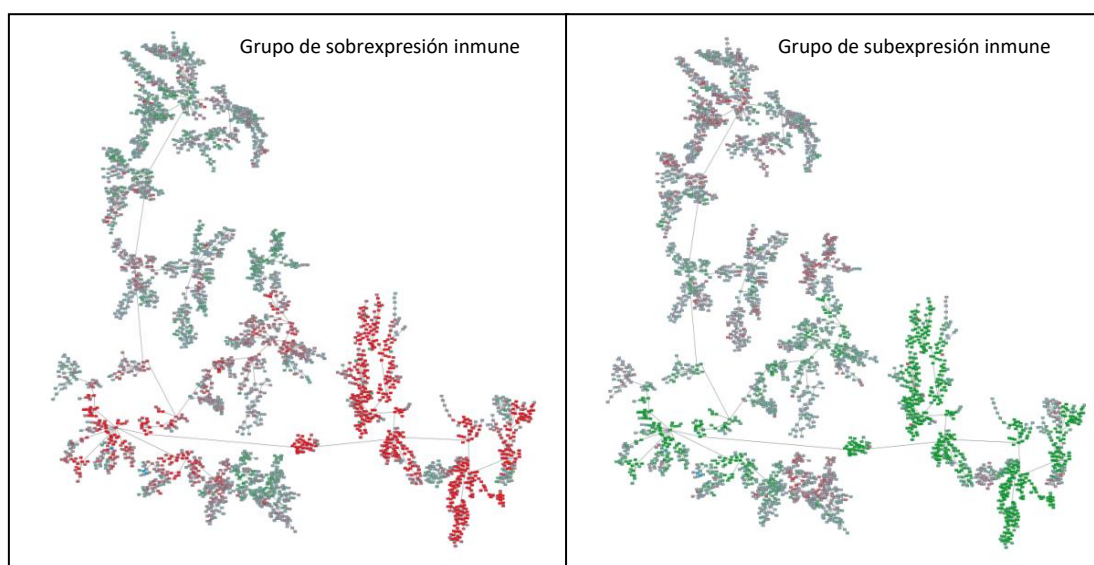


Figura 25. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. La expresión de los genes centrada a la media de expresión se representa en un mapa de calor que va de verde (-1) a rojo (+1) pasando por gris (0).

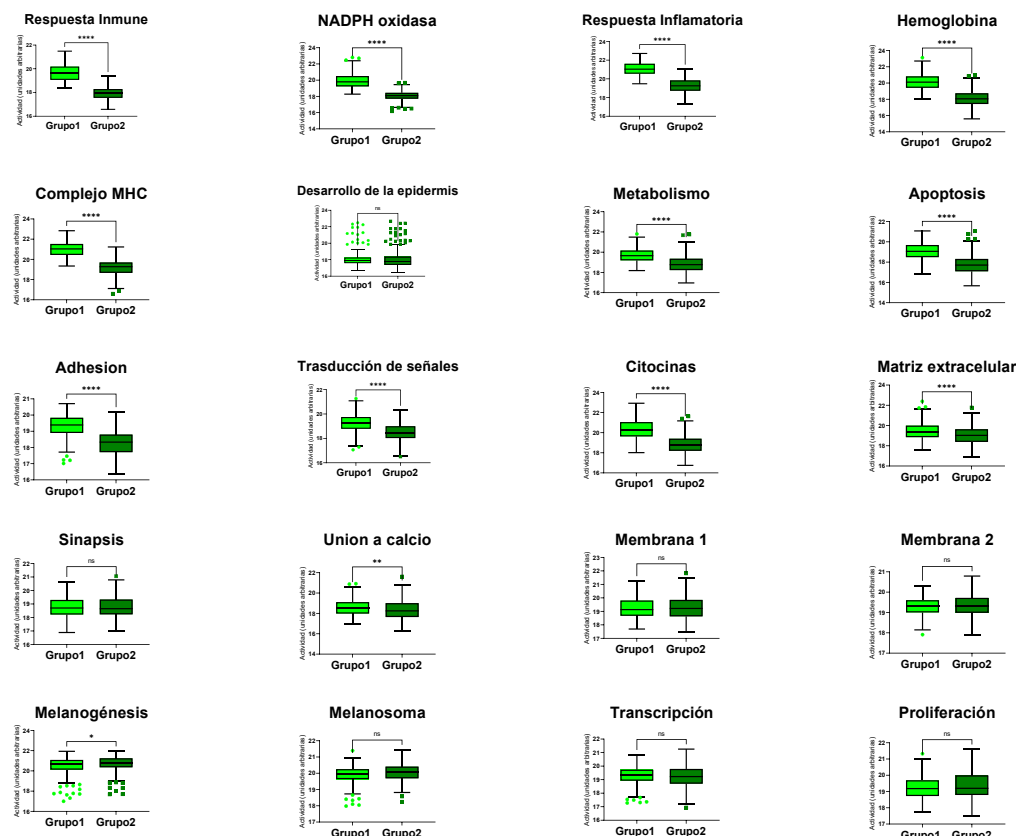


Figura 26. Actividad de los nodos del MGP en los grupos de la clasificación inmune.

1.4.2. Clasificación basada en información molecular del tumor

Las capas de información 2, 4, 5 y 7 mostraron en el HCL una posible clasificación en cuatro grupos. Utilizando los 280 genes de estas cuatro capas (figura 27) se realizó un CCA forzando a generar cuatro grupos: grupo 1 con 142 pacientes (30%), grupo 2 con 198 pacientes (42%), grupo 3 con 61 pacientes (13%) y grupo 4 con 71 pacientes (15%). Estos cuatro grupos se mantienen en un agrupamiento jerárquico con los 280 genes de las capas (figura 28).

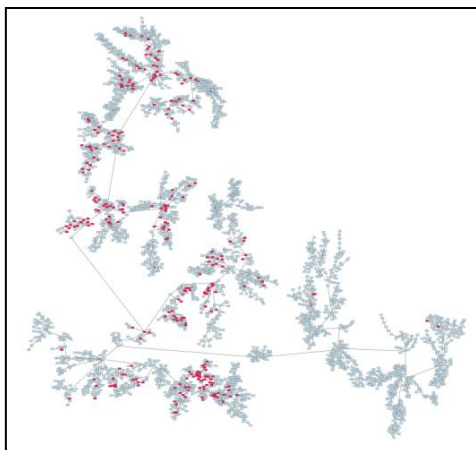


Figura 27. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. En rojo se representa los genes de las capas de información 2, 4, 5 y 7.

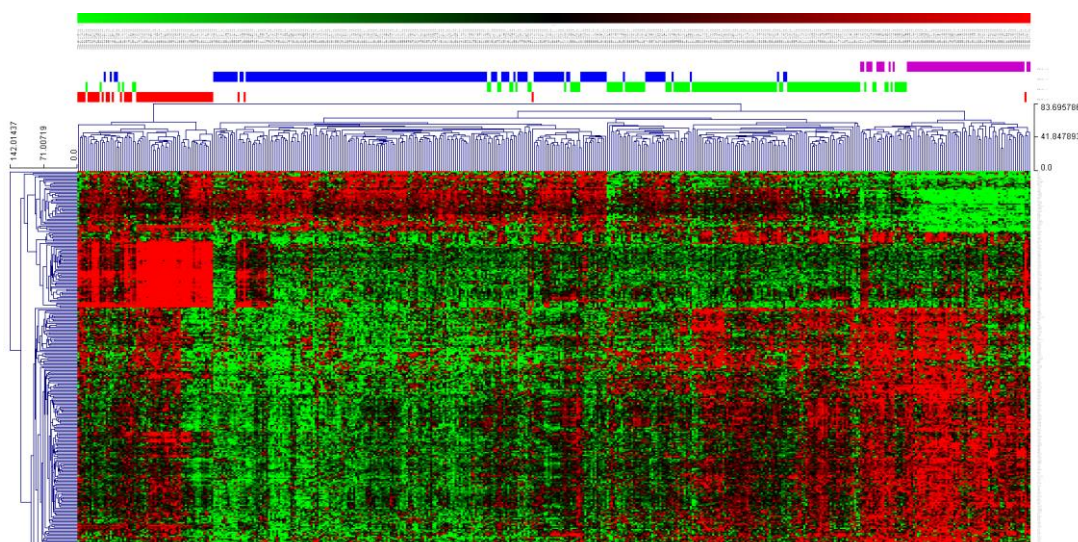


Figura 28. Agrupamiento jerárquico de la base de datos del TCGA. Se representa un HCL para los 472 pacientes (columnas) y otro para los 280 genes (filas). En colores están representados los grupos de la clasificación molecular. En verde los pacientes del grupo 1, en azul los pacientes del grupo 2, en rojo los pacientes del grupo 3 y en morado los del grupo 4.

Para caracterizar estos cuatro grupos se representó un mapa de calor en el MGP (figura 29). El primer grupo tiene sobreexpresión del nodo de melanosoma. El grupo 2 tiene sobreexpresión de los nodos de melanogénesis, proliferación y membrana 2. El grupo 3 tiene sobreexpresión del nodo de desarrollo de la epidermis y queratinas. Y el grupo 4 tiene sobreexpresión de los nodos adhesión, apoptosis, matriz extracelular, unión a calcio y membrana 1.

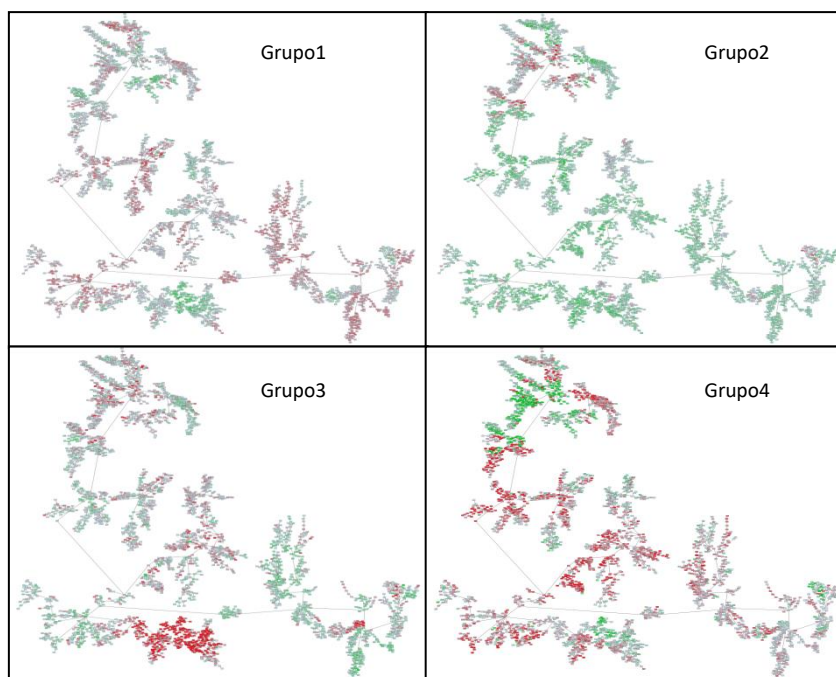


Figura 29. Modelo gráfico probabilístico de la expresión de los 2.971 genes más variables de la base de datos del TCGA. La expresión de los genes centrada a la media de expresión se representa en un mapa de calor que va de verde (-2) a rojo (+2) pasando por gris (0).

Para poder realizar una comparación entre las expresiones génicas de cada grupo se realizaron las actividades de los veintidós nodos del MGP (Figura 30). Destaca la mayor actividad del nodo de desarrollo de la epidermis y queratinas de las muestras del grupo 3. También se puede observar que el grupo uno tiene mayor actividad del nodo de melanosoma, el grupo 2 tiene mayor actividad del nodo de proliferación y el grupo 4 tiene mayor actividad del nodo de transcripción.

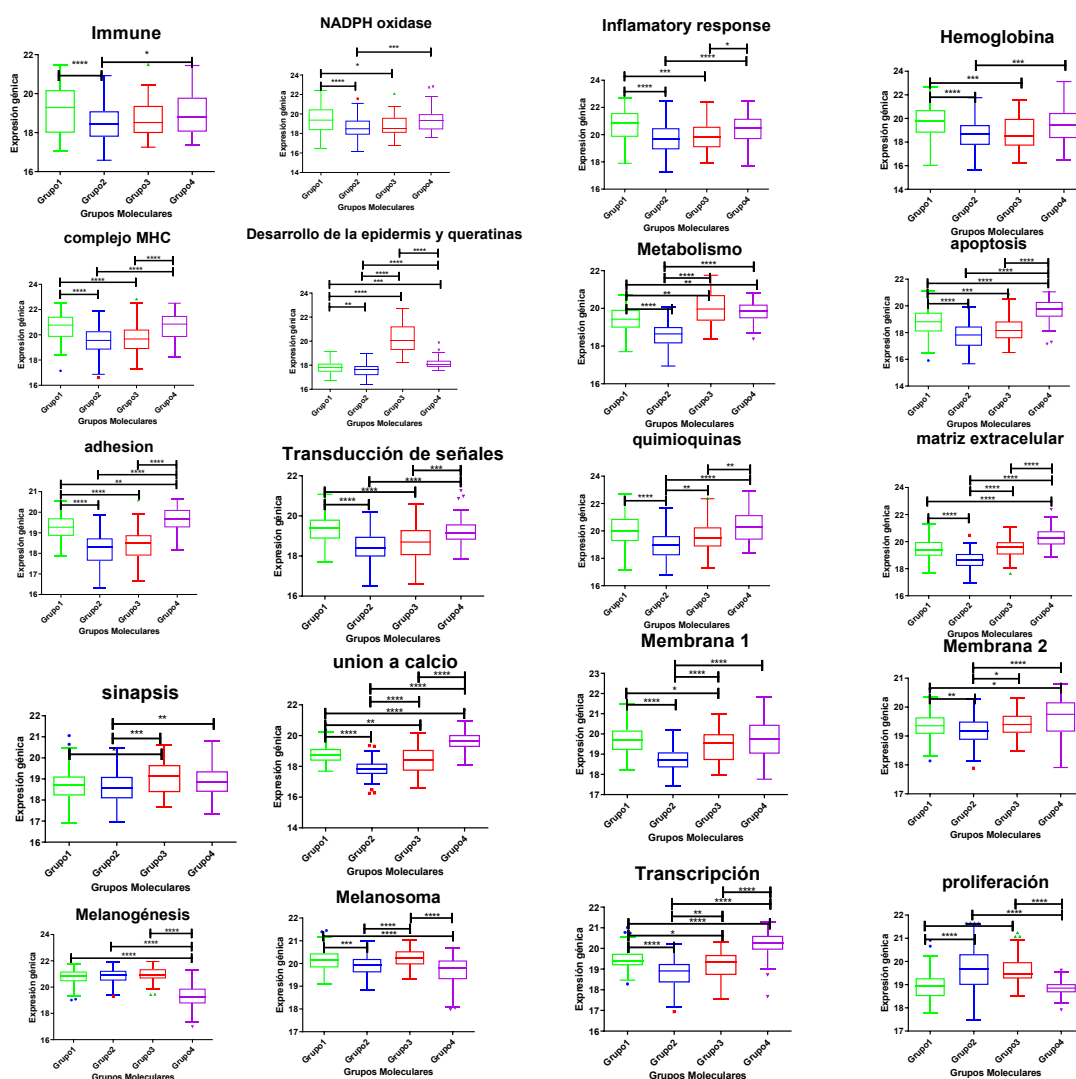


Figura 30. Actividad de los nodos de cada uno de los grupos moleculares.

El análisis de supervivencia muestra que no hay diferencias en la supervivencia libre de enfermedad de estos grupos (figura 31).

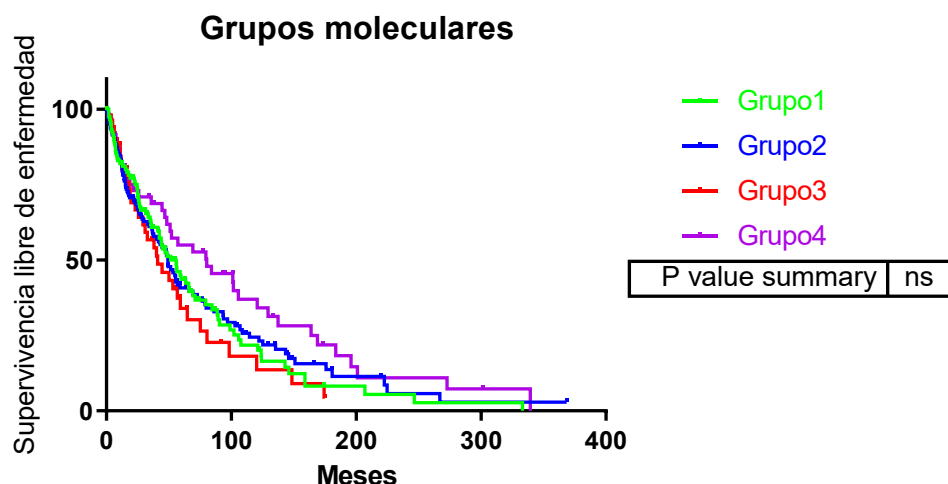


Figura 31. Curvas de supervivencia libre de enfermedad de los grupos de la clasificación molecular.

Los cuatro grupos moleculares pueden dividirse en función de la clasificación inmune (tabla 4).

Clasificaciones		Inmune	
		Alta expresión	Baja expresión
Molecular	Grupo 1	92 (65%)	50 (35%)
	Grupo 2	77 (39%)	119 (61%)
	Grupo 3	19 (31%)	42 (69%)
	Grupo 4	44 (62%)	27 (38%)

Tabla 4. Relación entre los grupos de la clasificación inmune y la clasificación molecular.

1.5. Resumen

A lo largo de esta sección se han desglosado los resultados más importantes obtenidos con la cohorte de descubrimiento. Se plantean dos clasificaciones distintas, una clasificación inmune que divide a los pacientes en dos grupos en función de si tienen alta o baja expresión de los genes de los nodos inmunes y una clasificación molecular que divide a los pacientes en cuatro grupos. Para llegar a estas clasificaciones independientes ha sido necesario analizar siete capas de información, cada una de las cuales tiene una serie de genes únicos y divide a los pacientes de la base de datos de una manera particular.

2. Cohorte de validación experimental del grupo español multidisciplinar del melanoma

De todos los resultados obtenidos en el análisis de la cohorte del TCGA, el resultado más relevante desde el punto de vista del posible impacto clínico es la información inmune, ya que podría estar relacionada con la respuesta al tratamiento con inmunoterapia. Sin embargo, la cohorte del TCGA no permite verificar esta hipótesis, puesto que es una serie previa a la llegada de la inmunoterapia moderna. Para evaluar la posible utilidad clínica de estos resultados, se empleó la cohorte de pacientes con melanoma avanzado recogida por el grupo español multidisciplinar del melanoma (GEM) del 2016 al 2019 que incluye pacientes que han recibido tratamiento inmunoterápico con diferentes anticuerpos anti-PD1.

2.1. Clínicos descriptivos

Se obtuvieron 53 muestras pertenecientes a 52 pacientes (uno de los pacientes tenía dos bloques) tratados con inmunoterapia (anti-PD1). Seis muestras se perdieron al cortar porque los bloques estaban muy gastados y no se pudieron obtener los cortes. Se añadió una muestra del Hospital Universitario La Paz que cumplía los criterios de inclusión.

Tras la extracción de ARN y antes del paso de diseño de las librerías se descartaron 8 muestras porque no había suficiente cantidad de ARN.

Por tanto, al final se utilizaron 40 muestras. La diferencia entre el número de muestras obtenidas y analizadas está recogida en el flujo de trabajo siguiente (figura 32).

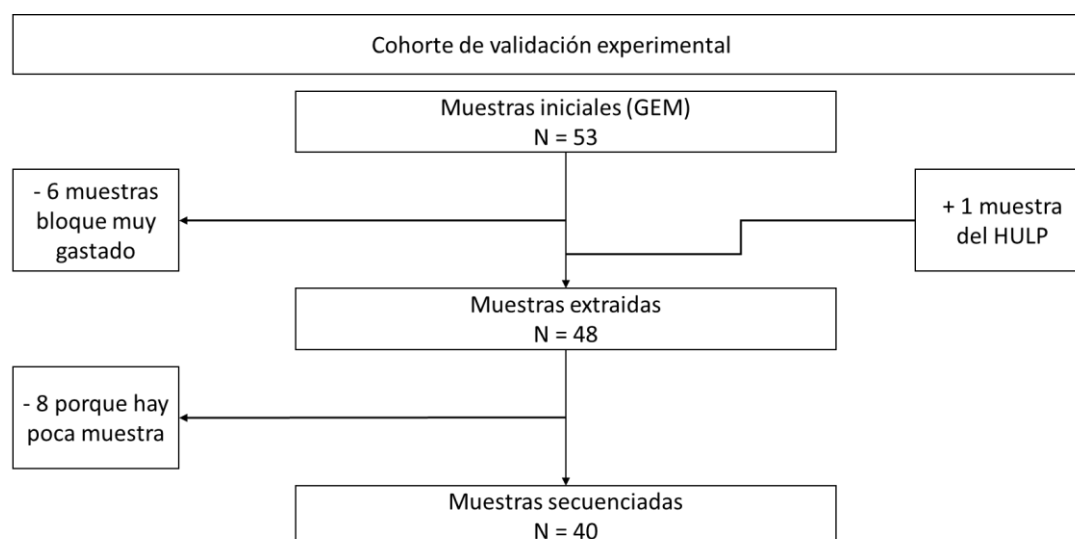


Figura 32. Flujo de trabajo del RNA-seq realizado en la cohorte de validación experimental.

De estas 40 muestras tenemos información clínica de 38 pacientes (porque un paciente tiene dos bloques y de otro no hay parámetros clínicos), presentadas en la tabla 5.

Características del paciente		Número	%
Género	Femenino	10	26%
	Masculino	28	74%
ECOG	Dato desconocido	6	16%
	0	27	71%
	1	5	13%
BRAF	WT	23	61%
	Mutado	10	26%
Tipo de melanoma	Origen desconocido	1	3%
	Cutáneo	35	92%
	Mucosas	2	5%
Tipo de muestra	Primario	20	53%
	Ganglio	5	13%
	Metástasis	13	34%
LDH	Elevado	8	21%
	Normal	30	79%
Metástasis	1-2	26	68%
	>2	12	32%
Línea de tratamiento	1	31	82%
	2	4	11%
	3	3	8%
Tratamiento Anti-PD1	Nivolumab	16	42%
	Pembrolizumab	22	58%
Mejor Respuesta a Anti-PD1	Respuesta completa	7	18%
	Respuesta parcial	10	26%
	Enfermedad estable	7	18%
	Progresión	10	26%
	No evaluable	3	8%
Tratamiento hiperprogresión	Si	4	11%
	No	31	82%
	Desconocido	3	8%

Tabla 5. Características de los pacientes de la base de datos de validación experimental. La columna número hace referencia al número de pacientes con cada característica y la columna porcentaje hace referencia al porcentaje de pacientes de cada característica en relación con el total de pacientes de la base de datos de validación experimental.

2.2. Análisis de expresión génica por RNA-seq

Primero se diseñó la captura empleando la plataforma informática *Nimbledesign* (Roche), de manera que no superase las 7 Megabases, es decir una captura de 1.964 genes que estaban identificados según el *gene symbol* oficial. Debido a que esta nomenclatura varía en función de la base de datos de referencia que se use se decidió traducir los *gene symbol* a la nomenclatura oficial de la base de datos ensambl. En este paso se perdieron

5 genes que no estaban catalogados en Ensembl y, por tanto, tenemos una captura de 1.959 genes. Estos 1.959 genes tienen dos orígenes distintos. Quisimos testar una serie de 260 biomarcadores de sistema inmune seleccionados a partir de un estudio bibliográfico realizado en el laboratorio (Angelova *et al.*, 2015). Además, queríamos introducir el mayor número de genes posibles de la base de datos de descubrimiento. De los 260 biomarcadores, 185 ya estaban presentes en la red de descubrimiento. De esta manera, la captura final se compuso de 1.885 genes de la base de descubrimiento (de los cuales 185 son marcadores de sistema inmune) más 75 biomarcadores que introdujimos nuevos (figura 33).

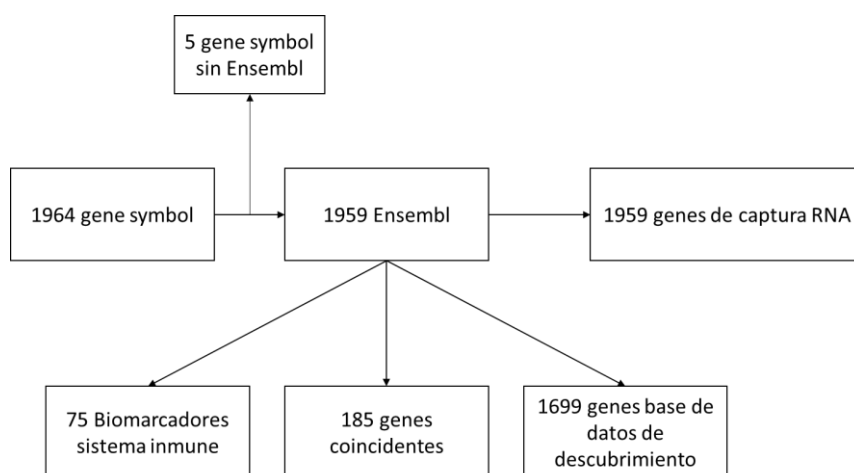


Figura 33. Diseño de la captura de ARN.

Una vez diseñada la captura se realizó la ligación y el producto resultante se amplió por PCR. En este caso se realizaron 5 pools de librerías, de 8 en 8 muestras, y después se realizaron capturas de cada uno de esos pools a partir de 1µg de cada pool.

2.3. Resultados RNA-seq

Una vez realizadas las capturas de ARN se realizó la secuenciación de las muestras en dos grupos de veinte. Se secuenciaron por tanto las 40 muestras y se obtuvieron un total de 484.692.023 lecturas.

Se creó una secuencia de comandos en R para sacar los valores de expresión de los 1.960 genes preseleccionados en la captura de ARN de entre todas las lecturas (material suplementario 1). 1.828 genes pasaron al criterio de selección (más de un 50% de los pacientes presentaban expresión para el gen por encima del umbral de detección), lo que supone un 93% de genes que pasan el criterio.

2.4. Modelo gráfico probabilístico

Con estos 1.828 genes se realizó un modelo gráfico probabilístico. La ontología génica de cada una de las ramas del MGP resultante fue analizada y dio como resultado el modelo de la figura 34, que consta de 16 nodos funcionales.

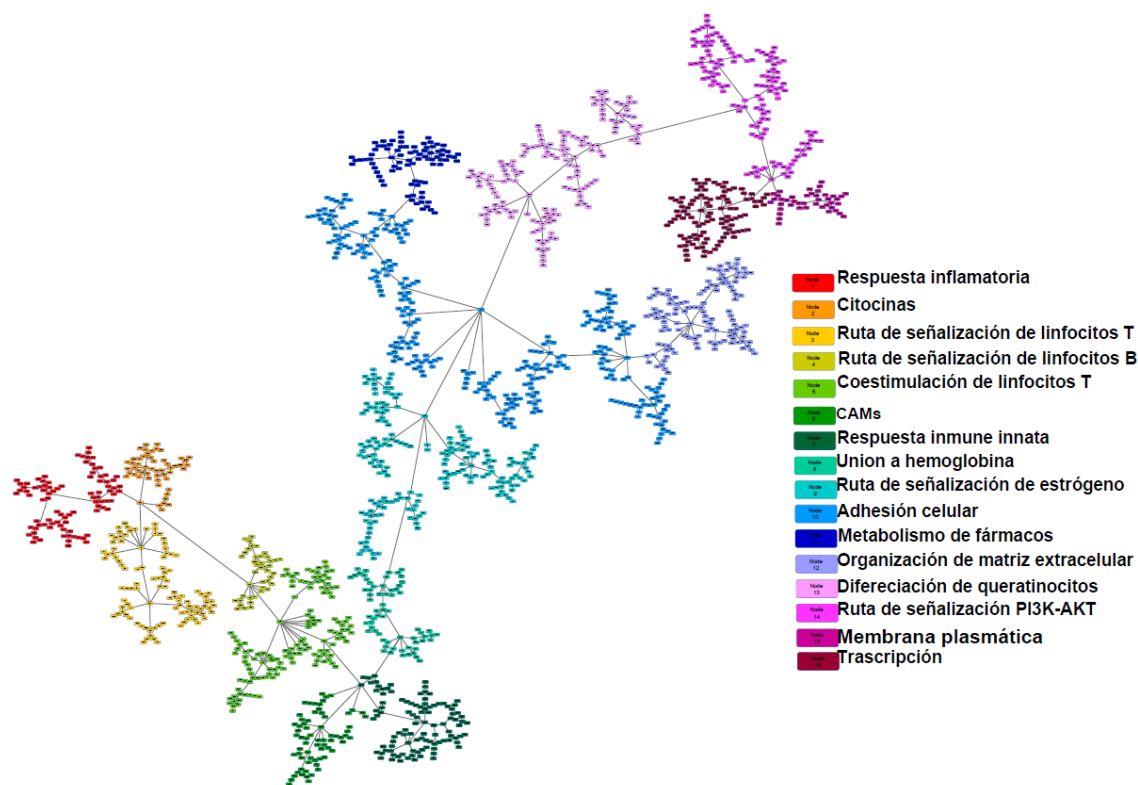


Figura 34. MGP a partir de los 1.828 genes de la base de datos del GEM. Cada nodo funcional tiene un color diferente comprobable en la leyenda de la parte derecha de la figura.

2.5. Capas de información de la base de datos de descubrimiento

Los genes de cada una de las capas de información de la base de datos del TCGA están representados en su mayoría en la base de datos del GEM, lo cual permitió trasladar la información de una base de datos a otra.

Utilizando el algoritmo CCA se obtuvieron grupos de muestras para cada una de las siete capas de información de la base de datos de descubrimiento (tabla 6).

Capa de información	Genes	% genes	Ontología génica	Nº grupos	% muestras grupo 1/2/3*
1	106	95%	Respuesta Inflamatoria	2	70%/30%
2	52	93%	biosíntesis de Melanina	3	37,5%/50%/12,5%
3	78	52%	Respuesta Inmune	2	50%/50%
4	62	58%	Ruta de señalización PI3k-AKT	2	47,5%/52,5%
5	46	96%	Desarrollo de la epidermis	2	57,5%/42,5%
6	54	42%	Respuesta Inmune	2	27,5%/72,5%
7	37	53%	Homeostasis del ion calcio	2	60%/40%

Tabla 6. Distribución de los grupos de las capas de información en la base de datos del GEM.

2.6. Supervivencia de los grupos de las capas

Se realizaron análisis de supervivencia para cada uno de los grupos de las capas y se obtuvo el siguiente resultado (figura 35).

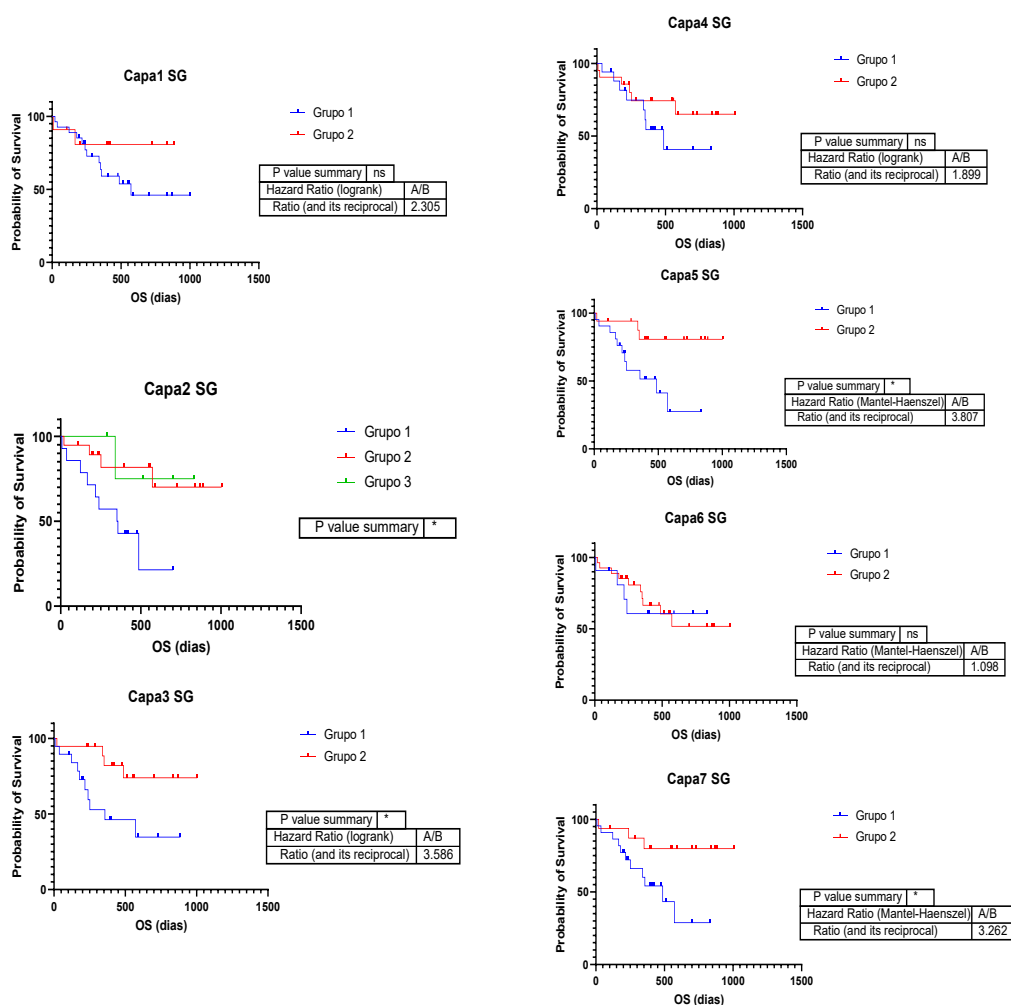


Figura 35. Supervivencia global de los grupos de pacientes de cada una de las capas de la base de datos de descubrimiento. * $p < 0,05$.

Las capas 2, 3, 5 y 7 dividieron a los pacientes de la cohorte en dos o más grupos con diferente supervivencia global. A continuación, se desglosarán cada una de estas capas.

2.6.1. Capa 2. Biosíntesis de melanina

La segunda capa de información está compuesta por 52 de los 56 genes de la capa del TCGA, un 93%. La ontología génica principal es biosíntesis de melanina, y los genes se distribuyen de la siguiente manera en el MGP (figura 36).

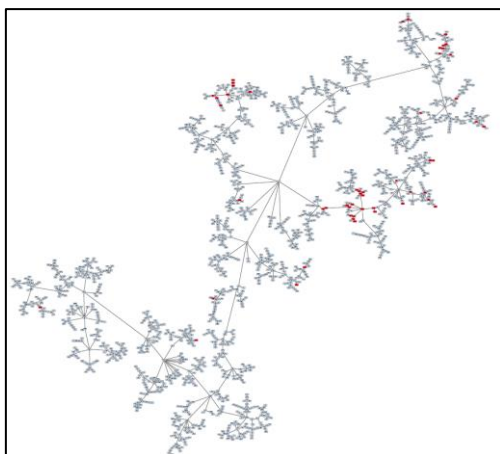


Figura 36. Distribución de los genes de la capa de biosíntesis de melanina en el modelo gráfico probabilístico.

Los pacientes se dividieron en 3 grupos según el CCA. El grupo 1 presenta un peor pronóstico. Este grupo destaca por tener una mayor expresión del nodo de queratinización (figura 37). El segundo grupo muestra una menor expresión en el conjunto de la red. Además, se identificó un tercer grupo que parece tener gran expresión los genes del nodo de las citocinas.

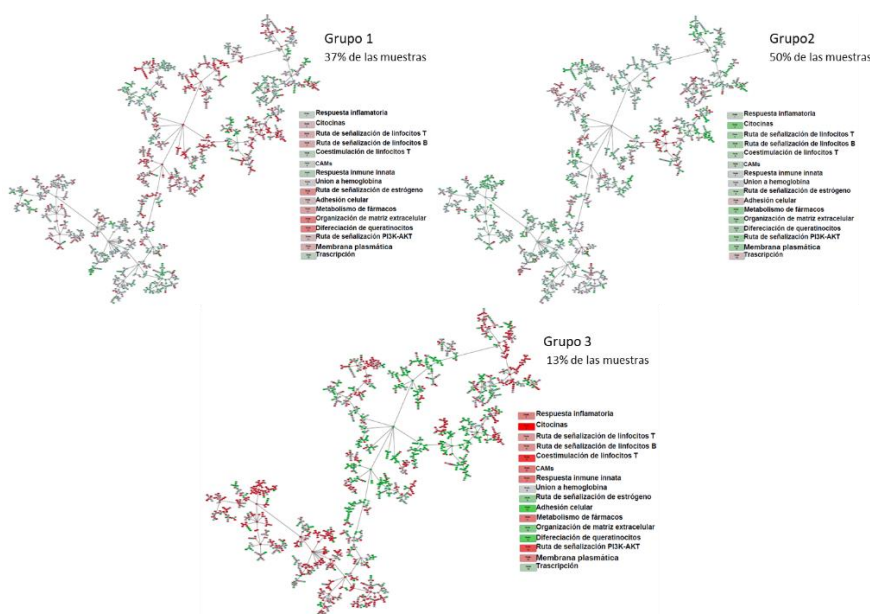


Figura 37. Expresión de los genes de la red del GEM en cada uno de los tres grupos de la capa de biosíntesis de melanina.

2.6.2. Capa 3. Respuesta inmune

La tercera capa de información, cuya ontología principal es respuesta inmune, está compuesta por 78 genes de los 150 que había en la base de datos del TCGA (52%). Estos genes están en su mayoría en los nodos de receptor de citocinas, Células T, células B y respuesta inmune innata (figura 38).

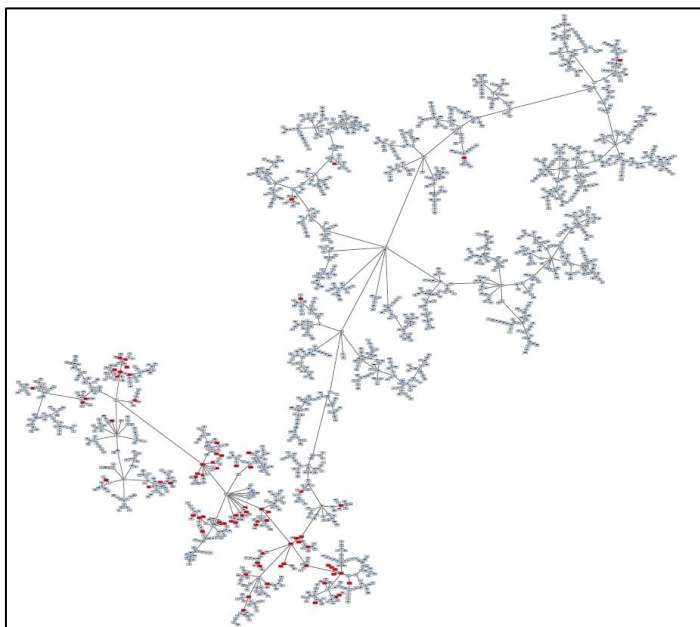


Figura 38. MGP con los genes de la capa 3 destacados.

Los pacientes se dividieron en dos grupos según el CCA con diferente expresión de los nodos del MGP. El grupo de peor pronóstico tiene menor expresión de los nodos de receptor de citocinas, Células T, células B y respuesta inmune innata. El grupo de mejor pronóstico tiene mayor expresión de estos nodos (figura 39).

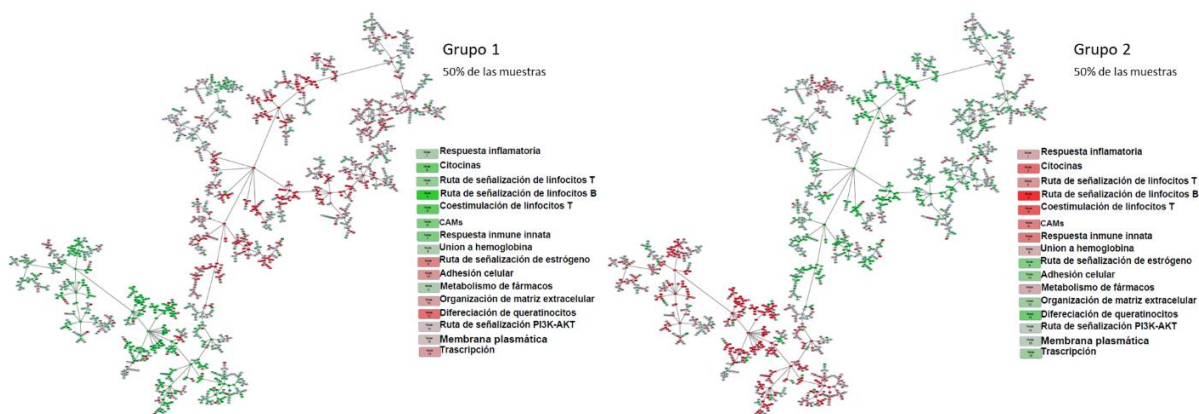


Figura 39. Mapa de calor de expresión de los genes del MGP para cada uno de los grupos de la capa 3.

2.6.3. Capa 5. Desarrollo de la epidermis

La quinta capa de información está compuesta por 46 de los 48 genes de la capa del TCGA un 96%. La ontología génica principal es desarrollo de la epidermis, y los genes se distribuyen de la siguiente manera en el MGP (figura 40).

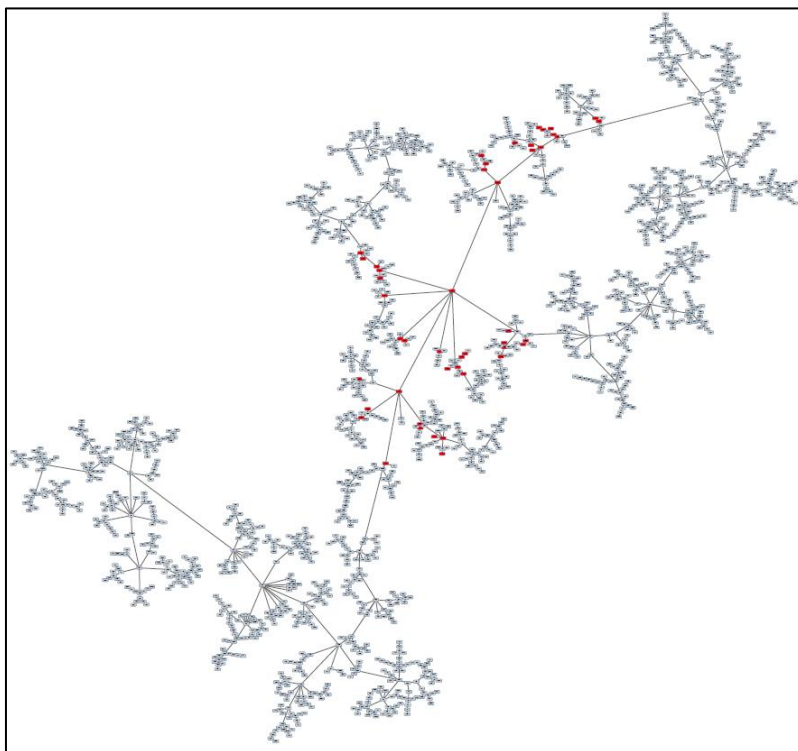


Figura 40. Distribución de los genes de la capa 5 en el MGP de la base de datos de validación experimental.

Los pacientes se dividieron en 2 grupos según el CCA con diferente expresión de los nodos del MGP. El grupo de peor pronóstico tiene mayor expresión del nodo queratinización. El grupo de mejor pronóstico tiene menor expresión de este nodo (figura 41).

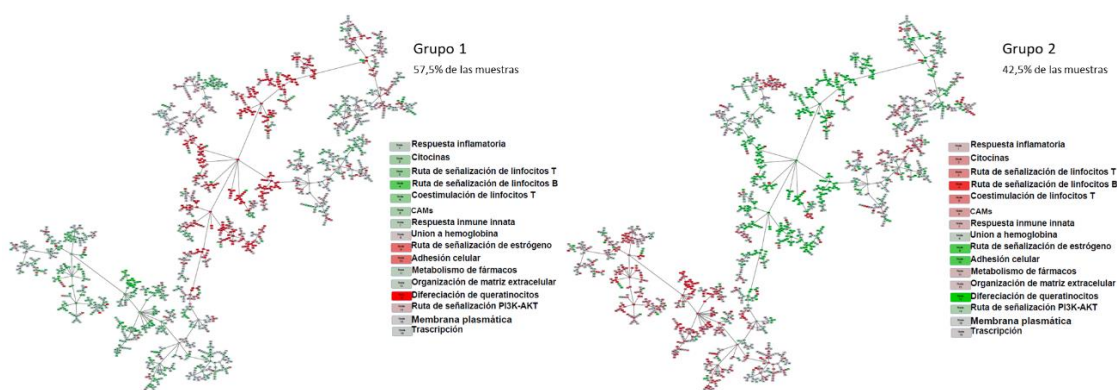


Figura 41. Mapa de calor de expresión de los genes del MGP para cada uno de los grupos de la capa 5.

2.6.4. Capa 7. Homeostasis del ion Calcio

La séptima capa de información está compuesta por 37 de los 70 genes de la capa del TCGA un 53%. La ontología génica principal es Homeostasis del ion Calcio, y los genes se distribuyen de la siguiente manera en el MGP (figura 42).

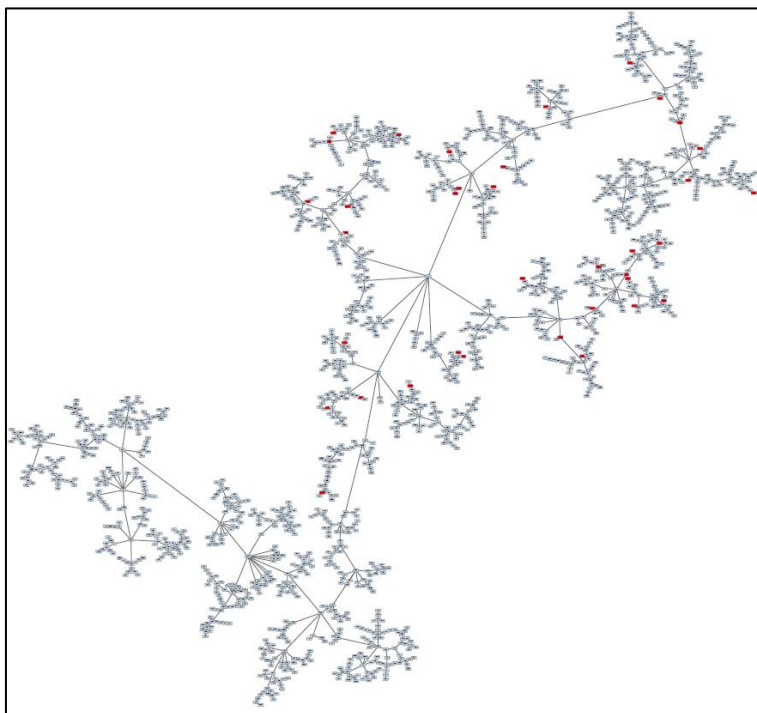


Figura 42. Distribución de los genes de la capa 5 en el MGP de la base de datos del GEM.

Los pacientes se dividieron en 2 grupos según el CCA con diferente expresión de los nodos del MGP. El grupo de peor pronóstico tiene mayor expresión de los nodos de queratinización, ruta de señalización del receptor de estrógenos y adhesión celular. El grupo de mejor pronóstico tiene menor expresión de los mismos nodos (figura 43).

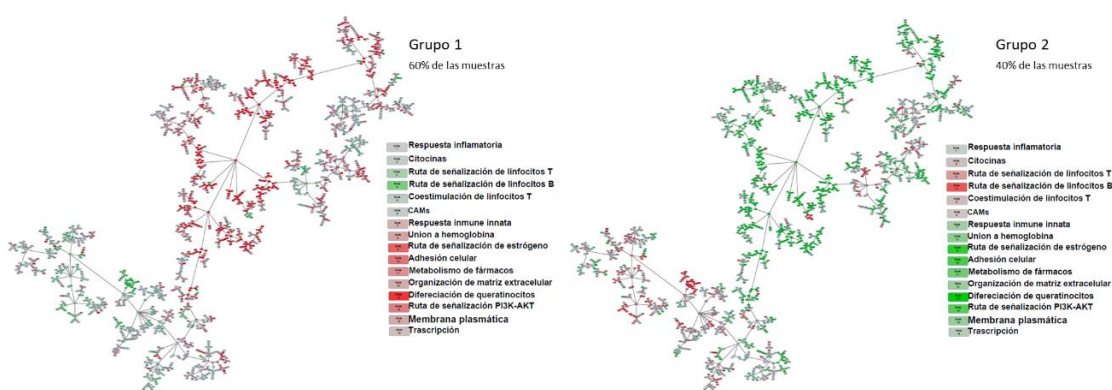


Figura 43. Mapa de calor de los grupos de la capa 7.

2.6.5. Resumen de las capas

La clasificación del CCA de los pacientes en las capas de información 2, 5 y 7 identifica un grupo con peor pronóstico que tiene mayor expresión del nodo de queratinización. Además de estas capas, la única que tiene valor pronóstico es la capa 3, cuya ontología principal es respuesta inmune. Esto confirma las tendencias observadas en la cohorte del TCGA, pero, debido al bajo número de muestras las diferencias entre las actividades de los nodos en los grupos no son significativas.

DISCUSIÓN

Discusión

1. Hallazgos de la tesis

El análisis por medio de herramientas bioinformáticas de dos cohortes de pacientes con melanoma ha resultado en la propuesta de diversas clasificaciones que atienden a la realidad del tumor. El primer hallazgo fundamental del trabajo es que los melanomas pueden dividirse en función de su estatus inmune y de una serie de marcadores moleculares, y que ambas clasificaciones funcionan independientemente. Un segundo hallazgo relevante de esta tesis doctoral es que, en una cohorte de pacientes de melanoma avanzado tratados con inmunoterapia, el análisis a través de capas de información identifica grupos con distinta respuesta a dicha inmunoterapia.

2. Antecedentes del trabajo

La hipótesis de esta tesis surge a raíz del artículo publicado en Scientific Reports donde se realizó un análisis molecular del cáncer de mama triple negativo (Prado-Vázquez *et al.*, 2019). En este artículo se aseveraba la existencia de dos tipos de información biológica, la relativa al sistema inmune del paciente y la que tenía que ver de manera más directa con las células tumorales, y que estas informaciones funcionan de forma independiente. El primer objetivo de la tesis fue la búsqueda de estos tipos de información en un tipo de cáncer distinto como es el melanoma.

3. Puesta en valor de la metodología de trabajo cohorte publica descubrimiento-cohorte experimental de validación

Para ello se descargó la base de datos del consorcio The Cancer Genome Atlas (TCGA) porque de las bases de datos accesibles evaluadas era la más completa. Dispone de 469 muestras analizadas por ARN-seq, para cada una de las cuales tiene recogidos también los datos clínicos (Tomczak, Czerwińska and Wiznerowicz, 2015). El uso de bases de datos públicas como las del TCGA facilita la generación de hipótesis previas a un estudio experimental en una nueva cohorte prospectiva o retrospectiva. Indudablemente, los resultados obtenidos en estas investigaciones *in silico* deben de ser comprobados en una serie de muestras supervisadas desde el inicio para poder confirmar cualquier hipótesis. Sin embargo, la generación de hipótesis con este método es tiene varias ventajas. En primer lugar, no hay que esperar a obtener las muestras y realizar la extracción de ARN de estas, sino que directamente se comienza con el análisis de datos. Además, estas cohortes suelen incluir un gran número de pacientes. En contra, suelen presentar una gran heterogeneidad en los pacientes incluidos, y la información clínica suele ser escasa e incompleta.

Por tanto, en este trabajo se usó la base de datos del TCGA como una base de datos de descubrimiento y, una vez definida una hipótesis (que la información de las capas inmunes puede tener valor clínico relacionado con la respuesta a la inmunoterapia), se analizaron pacientes de una segunda cohorte obtenida gracias a la colaboración con el Grupo Español Multidisciplinar de Melanoma (GEM), que sirvió de base de datos de validación.

4. Puntualizaciones técnicas

Para el análisis de la cohorte de validación se emplearon dos técnicas novedosas en el análisis de datos de origen ómico: los modelos gráficos probabilísticos, y el *sparse k-means*/CCA de forma recursiva. Antes de discutir los resultados proporcionados por ambas técnicas, conviene hacer una serie de puntualizaciones. Tras varios trabajos en el laboratorio utilizando los modelos gráficos probabilísticos, se ha llegado a la conclusión de que lo óptimo es emplear entre 1000. y 3.000 genes con la mayor varianza en la serie (Trilla-Fuertes *et al.*, 2019b; Zapater-Moros *et al.*, 2018; Gámez-Pozo *et al.*, 2017b; Prado-Vázquez *et al.*, 2019). Los genes con mayor varianza reflejan las diferencias entre las muestras de la base de datos y son los que tienen más potencial a la hora de caracterizar los tumores y buscar subtipos. Es posible que de esta manera se excluyan genes característicos del tipo de tumor estudiado porque todas las muestras tengan mucha expresión de dicho gen. Sin embargo, en anteriores estudios hemos llegado a la conclusión que este caso es poco probable y, en caso de ocurrir, no debería afectar al objetivo del estudio, que es encontrar diferencias entre tumores de un tipo específico de cáncer.

5. Caracterización del MGP

Los MGPs son una representación gráfica de la base de datos utilizada. La clave de estos modelos es que funcionan sin ningún apriorismo, es decir, que organizan los genes en función de su expresión a lo largo de una serie sin que previamente se seleccione ningún tipo de variable en base a la hipótesis (Gámez-Pozo *et al.*, 2015). Por tanto, si dos genes se sitúan contiguos en la red es porque la expresión de ambos en la base de datos es similar. El MGP de la base de datos de descubrimiento, por ejemplo, se dividió en 21 nodos funcionales. Esto significa que hay 21 bloques de genes con una expresión similar en la base de datos del TCGA, que además tienen una ontología génica relacionada. Este es uno de los primeros resultados del trabajo, pues demuestra que los genes con una misma función se relacionan en la red.

5.1 Similitudes y diferencias con MGP de trabajos previos

Un estudio previo de nuestro grupo con MGP en cáncer de mama triple negativo describió 26 nodos funcionales, entre los cuales destacaban nodos de respuesta inmune, de adhesión celular, de queratinización o de membrana plasmática, nodos que también presentan los MGPs aquí presentados (Prado-Vázquez *et al.*, 2019). Los MGP desarrollados en otros tipos de cáncer, como el de vejiga músculo invasivo, describen los mismos nodos (Trilla-Fuertes *et al.*, 2019b; de Velasco *et al.*, 2017). Sin embargo, el nodo de melanogénesis solo está presente en los MGPs de los estudios realizados con melanomas, ya sean basados en expresión génica o en proteómica (Trilla-Fuertes *et al.*, 2019c).

Por tanto, los MGP captan una serie de procesos tumorigénicos generales que se pueden asociar a las características primordiales (“hallmarks”) del cáncer descritas por Hanahan y Weinberg; la evasión de la destrucción inmunológica (nodos de respuesta inmune), la activación de la invasión y la metástasis (adhesión celular) e inducción de la angiogénesis (nodos de membrana plasmática) (Hanahan and Weinberg, 2011). Pero también recogen procesos moleculares relevantes propios de cada tipo de tumor, como la melanogénesis en los MGP de melanoma.

5.2 Procesos interesantes detectados en los nodos

Uno de los principales retos que supone el trabajar con MGP es priorizar en que información es más adecuado profundizar, ya que un estudio de estas características da lugar a un gran número de hipótesis.

A continuación, vamos a desarrollar los hallazgos de los nodos de Respuesta inmune, respuesta inflamatoria, complejo MHC, NADPH oxidasa, desarrollo de epidermis, melanogénesis y melanosoma.

5.2.1 Inmunes

Cabe reseñar que los nodos que tienen que ver con el sistema inmune (inmune, respuesta inflamatoria, complejo MHC) están unidos en un grupo en un extremo de la red (figura 6). Los principales biomarcadores de respuesta inmune hasta la fecha estudiados en melanoma son PD-1 (*PDCD1*), PDL-1 (*CD274*) y CTLA-4 (*CTLA4*). Estos marcadores son también dianas de la inmunoterapia, pero al menos en el campo del melanoma no han servido para seleccionar a los pacientes con más probabilidad de responder a dicho tratamiento (Wolchok *et al.*, 2017; Robert *et al.*, 2019). Por ejemplo, PD-L1 sólo se correlaciona débilmente con la respuesta en el melanoma avanzado (Johnson and Sosman,

2015). Por otro lado, la expresión de *MHC-II* (nodo del complejo MHC) en células tumorales se ha validado en varios estudios (Daud *et al.*, 2016). Nosotros proponemos una aproximación por nodos en la que se estudiase no solo la expresión de uno de estos genes, sino la de todos los genes de su nodo, que tienen una expresión similar en la base de datos del TCGA.

5.2.2 NADPH oxidasa

En cuanto al nodo de NADPH oxidasa, varios trabajos sugieren que el melanoma es un tumor dirigido por la presencia de especies reactivas de oxígeno (ROS), y que este proceso puede estar relacionado con la génesis y evolución de melanoma, así como con la resistencia a fármacos como el vemurafenib (Fried and Arbiser, 2008; Meyskens *et al.*, 2007; Wittgen and van Kempen, 2007; Liu-Smith and Meyskens, 2016). Por otro lado, la presencia de ROS se ha relacionado previamente con el proceso de síntesis de la melanina (Obrador EObrador *et al.*, 2019), si bien ambos procesos se encuentran distanciados en red generada. Resulta sorprendente que el nodo de NADPH oxidasa se encuentre inmerso entre los nodos inmunes. Hay algunos trabajos previos que relacionan la presencia de óxido nítrico en el microambiente tumoral (TME) con distintos efectos sobre el sistema inmune. Se ha sugerido que niveles elevados de NO en el TME del melanoma ayuda a mantener un estado inflamatorio crónico, que a su vez mantiene el TME en un microambiente inmunosupresor (Tanese, Grimm and Ekmekcioglu, 2012), ya que el aumento de los niveles de NO puede conducir a una menor proliferación de leucocitos y una menor infiltración (Choudhari *et al.*, 2013). Esto se refleja también en nuestros resultados, puesto que los nodos NADPH oxidasa y respuesta inflamatoria están contiguos. En cuanto a su relación con la respuesta a terapia, en un estudio preliminar en ratones, el tratamiento combinado con un inhibidor de NOS y un anticuerpo anti-PD1 mostró mejores tasas de respuesta que los agentes individuales en tumores de mama triples negativos (Davila-Gonzalez *et al.*, 2017). Sin embargo, no hay ningún estudio mostrando este efecto en melanoma. Nuestros resultados, apoyado por estas evidencias previas, sugieren una relación entre el estatus inmune y la presencia de ROS. Son necesarios estudios específicos para dilucidar este posible el papel de ROS en la eficacia de la inmunoterapia, ya sea como factor de resistencia, o como una posible diana terapéutica para tratamientos de combinación.

5.2.3 Desarrollo de la epidermis y queratinización

Otro nodo destacable es el de desarrollo de epidermis y queratinas. En diversos estudios transcriptómicos de melanoma se presentan evidencias de la existencia de una expresión diferencial de las queratinas dependiendo de la cohorte (Winnepeninckx *et al.*, 2006;

Wang, Li and Chen, 2018; Jaeger *et al.*, 2007; Tímár, Gyorffy and Rásó, 2010). Un trabajo reciente comparó la expresión de 31 melanomas primarios y 73 metástasis y encontró 426 genes con expresión diferencial entre ambos grupos (Wang, Qu and Chen, 2019). Entre los 20 genes con mayor diferencia de expresión estaban la acuaporina 3 (*AQP3*), la envoplaquina (*EVPL*) y la queratina 14 (*KRT14*), que aparecen en el centro del nodo de desarrollo de epidermis y queratinas en nuestra red. La *AQP3* cumple un rol fundamental en el desarrollo de la epidermis y en la diferenciación de queratinocitos (Nakahigashi *et al.*, 2011). En concreto se ha visto que la sobreexpresión de *AQP3* aumenta la progresión del melanoma y la resistencia a fármacos (Gao *et al.*, 2012). Además se demostró que el transporte de agua oxigenada por la *AQP3* probablemente desata señales intracelulares que inician la progresión del cáncer (Satooka and Hara-Chikuma, 2016). Por otro lado, las queratinas son las principales proteínas estructurales de las células epiteliales (Gigi *et al.*, 1982). La expresión del par *KRT5 / KRT14* está estrechamente relacionada con la transición epitelial-mesenquimatosa (EMT), que tiene un efecto significativo sobre la progresión y metástasis tumorales (Savagner, 2010). Cabe destacar que, en el nodo de desarrollo de epidermis, *KRT5* y *KRT14* se encuentran unidos por una arista, lo cual indica que el MGP detecta la íntima relación entre la expresión de estos dos genes.

5.2.4 Melanogénesis y melanosoma

También hay que destacar los nodos de melanogénesis y melanosoma, que están juntos en la parte superior de la red. El nodo de melanogénesis surge como una ramificación del gen central de melanosoma, el gen del premelanosoma (*PMEL*). Los melanocitos son células productoras de pigmentos, donde la melanina se genera como gránulos de pigmento, también conocidos como melanosomas (Videira, Moura and Magina, 2013). *PMEL* es la proteína responsable de la formación láminas fibrilares que sirven como plantilla sobre la que polimerizan las melaninas a medida que se sintetizan. Se ha visto que animales que carecen de expresión de *PMEL* muestran diversos grados de hipopigmentación e inviabilidad de las células pigmentarias (Watt *et al.*, 2013). En este nodo están presentes también genes que tienen que ver con la biogénesis (*MLANA*) y maduración del melanosoma (*SLC45 A2*) y con enzimas catalizadoras de la melanina como la dopacromo tautomerasa (*DCT*), la tirosinasa (*TYR*) o la melastatina (*TRPM1*) (de Vries *et al.*, 1999; Ainger *et al.*, 2014; Lai *et al.*, 2018; Guo, Carlson and Slominski, 2012). Entre los genes del nodo de melanogénesis también se encuentra el gen de la melanofilina (*MLPH*), una proteína que está involucrada en la distribución de los melanosomas por la piel (Orgaz *et al.*, 2011). Todos estos genes tienen una expresión regulada por el factor de transcripción AP-2 Alpha (*TFAP2A*), que también está presente en el nodo de melanogénesis (Seberg *et al.*, 2017). La presencia conjunta de todos estos genes en los nodos de

melanogénesis y melanosoma viene a confirmar el correcto funcionamiento del MGP, pues valida el resultado de la expresión de los genes de la base de datos con evidencias descritas en trabajos previos. Por último, hay que destacar la presencia del gen *TYRP1* en este nodo. Su sobreexpresión ha sido relacionada con un peor pronóstico en melanomas metastásicos (Journe *et al.*, 2011).

5.2.5 Otros nodos del MGP

El nodo de hemoglobina está localizado entre los nodos inmunes en el MGP. Se ha observado relación entre la anemia y la invasión de los ganglios linfáticos, la aparición de metástasis y un peor pronóstico en pacientes con melanoma (Tas and Erturk, 2018).

La presencia de un nodo de metabolismo es relevante, ya que en trabajos previos se propone que el metabolismo del melanoma es heterogéneo y adaptativo, es decir, los melanomas son capaces de adaptarse y utilizar diversos tipos de combustible para facilitar la progresión y la metástasis (Fischer *et al.*, 2018). Además, estudios recientes sugieren también que el metabolismo oxidativo tumoral puede ser una diana para mejorar la respuesta a la inmunoterapia (Najjar *et al.*, 2019).

El nodo de adhesión celular contenía varias integrinas (*ITGA4* e *ITGA8*), las cuales están asociadas con angiogénesis y metástasis órgano-específicas en el melanoma (Huang and Rofstad, 2018).

En cuanto a la relevancia del nodo de transducción de señales, se han estudiado numerosas dianas terapéuticas en la transducción de señales en el melanoma cutáneo (Uzdensky, Demyanenko and Bibov, 2013). A pesar de que los genes del nodo de transducción de señales no se estudiaron específicamente en esta revisión, es posible que contenga nuevas dianas terapéuticas.

También se ha identificado un nodo de citocinas. Estas moléculas se han probado como fármacos inmunoterápicos, pero con pobres resultados (Conlon, Miljkovic and Waldmann, 2019). El nodo de citocinas contiene el receptor de la IL7, la cual se administró en dos ensayos clínicos fase I sin respuestas objetivas (Sportès *et al.*, 2010; Rosenberg *et al.*, 2006). Sería interesante evaluar en el futuro si la actividad del nodo de citoquinas tiene alguna relación con la respuesta al tratamiento con estas moléculas.

El MGP presenta también un nodo de matriz extracelular (MEC), la MEC juega un papel importante en la regulación del microambiente tisular y en el mantenimiento de la ho-

meostasis celular y, sin embargo, no se dispone de biomarcadores de MEC para la predicción de la respuesta terapéutica a quimioterapia (Napoli *et al.*, 2020). El nodo el MGP podría arrojar luz en este aspecto.

El nodo de unión a calcio de la red contiene 32 isoformas de protocadherinas que podrían tener un papel en la progresión del melanoma, pues se ha visto que las protocadherinas tienen un papel importante en la migración celular y la invasión en cáncer gástrico (Chen *et al.*, 2017).

De la misma forma se podrían estudiar los nodos de membrana en búsqueda de genes que expresen proteínas de membrana que pudieran ser dianas terapéuticas (Qiu and Wang, 2008).

Por último hay que destacar la presencia del gen SOX9 en el nodo de proliferación, cuya sobreexpresión está relacionada con la presencia de metástasis en melanoma (Yang *et al.*, 2019).

5.2.6 Resumen del MGP

En resumen, el modelo gráfico probabilístico identifica veintiún nodos funcionales que reflejan procesos biológicos relevantes en la génesis y evolución del melanoma, así como procesos relacionados con la resistencia a algunos fármacos. Hay que poner en valor que con un solo modelo gráfico se obtiene información de un gran número de procesos. En concreto destacan nodos que tienen que ver con la melanogénesis, el desarrollo de la epidermis y el sistema inmune. En esta tesis se propone usar los nodos funcionales de dos formas, (1) como biomarcadores más robustos que los basados en un único gen y (2) como posibles conjuntos de dianas terapéuticas.

6. Las capas de información

6.1 Consideraciones al análisis de capas

En este trabajo aplicamos un nuevo sistema de clasificación por capas de información. La idea de este sistema es introducir el menor sesgo posible a la hora de realizar las clasificaciones de las muestras en las bases de datos. Para ello utilizamos dos algoritmos de clasificación. El primero es el *sparse k-means* (Witten and Tibshirani, 2010) y el segundo es el análisis de grupos consenso (*consensus cluster analysis*, CCA) (Monti *et al.*, 2003). En este caso se identificaron siete capas de información diferentes. Nos limitamos a siete capas de información porque se observó redundancia entre capas: la capa 3 y la 6 tenían que ver con respuesta inmune y la capa 7 y la capa 4 tenían gran similitud a la hora de dividir a las muestras de la base de datos. No obstante, sería posible hacer capas hasta

que se acabasen los 2.971 genes de la base de datos. Se podría ir incluso más lejos y hacer capas para el total de genes de la base de datos. Sin embargo, dudamos que esto aporte información adicional, ya que sabemos que en la identificación de las capas influye la variabilidad presente en la base de datos, del mismo modo que el MGP no proporciona más información al incluir genes con poca varianza.

El análisis por capas de información es un método que resulta especialmente útil a la hora de generar clasificaciones de muestras sin incluir ninguna información previa. Además, este análisis genera clasificaciones no excluyentes, es decir, una misma muestra puede pertenecer a tantos grupos como capas haya. Esto es muy importante para analizar todas las dimensiones de una enfermedad.

6.2 Reorganización de las capas en grupos

6.2.1 Racional de este estudio

La ontología principal de las capas estudiadas gira en torno a los dos tipos de información biológica de los que hablábamos al principio de la discusión: sistema inmune del paciente y biología de las células tumorales (Prado-Vázquez *et al.*, 2019). Las capas 1, 3 y 6 tienen que ver de manera más o menos directa con el sistema inmune, mientras que las capas 2, 4, 5 y 7 tienen que ver con la información molecular del melanoma.

6.2.1.1 Clasificación Inmune

El agrupamiento jerárquico (HCL) realizado en la base de datos del TCGA reflejó que las capas 1,3 y 6 dividían las muestras en dos grupos muy similares. A partir de este resultado se buscó la manera de agrupar estas clasificaciones. Para ello se eligieron los genes de las tres capas y se aplicó el CCA, que resultó en dos grupos: un grupo compuesto por las muestras con mayor expresión de los genes inmunes (Inmune positivo) y un segundo con menor expresión de estos genes (Inmune negativo). Se realizó un estudio de supervivencia libre de enfermedad comparando ambos grupos y no se encontraron diferencias. Desde los primeros estudios con inmunoterapias anti PD1 (Schadendorf *et al.*, 2015) ha aumentado la supervivencia de los pacientes con melanoma, pero la cohorte del TCGA es anterior a estos estudios, así que no contiene pacientes tratados con anti-PD1. Un estudio de una base de datos con pacientes de melanoma tratados con inmunoterapia podría arrojar luz a esta clasificación.

6.2.1.2 Clasificación Molecular

Utilizando las demás capas se creó una clasificación a la cual se denominó clasificación molecular que tenía cuatro grupos; un grupo de sobreexpresión de genes de desarrollo de

epidermis y queratinas, otro grupo con menor actividad del nodo de melanogénesis y melanosoma, otro con mayor actividad de los nodos de metabolismo, adhesión celular y apoptosis y un último grupo que solo tiene alta expresión del nodo de membrana. El grupo de desarrollo de epidermis y queratinas presenta un pronóstico ligeramente peor que el resto, aunque la diferencia no es significativa. Este resultado concuerda con lo presentado por el TCGA, donde identifican un grupo de pacientes con peor pronóstico que tienen alta expresión de las queratinas (The Cancer Genome Atlas Network, 2015). El grupo de menor actividad de los nodos de melanogénesis y melanosoma está relacionado con el grupo “MITF-low” del TCGA pues ambos presentan baja expresión de los genes asociados con pigmentación. El grupo de mayor actividad de los nodos de metabolismo, adhesión celular y apoptosis al igual que el grupo de alta expresión de genes de membrana sin embargo no se identificaron en la clasificación del TCGA (The Cancer Genome Atlas Network, 2015).

6.3 Comparación con la clasificación transcriptómica del TCGA

El consorcio del TCGA propuso una clasificación “transcriptómica” del melanoma (The Cancer Genome Atlas Network, 2015). Como ya hemos desglosado en la introducción, esta clasificación tiene tres grupos: “inmune”, “queratinas” y “MITF-low”. Si comparamos esta clasificación con las nuestras se puede ver que hay similitudes y diferencias. La diferencia fundamental es que nuestra aproximación estudia el estatus inmune de los pacientes de manera independiente. En la clasificación del TCGA, un 51% de los pacientes fueron catalogados como inmunes en base a un agrupamiento jerárquico, si bien estas muestras no expresaban de manera homogénea el resto de genes estudiados (The Cancer Genome Atlas Network, 2015). En nuestro trabajo separamos la clasificación inmune, de manera que un 49% de los pacientes se catalogan como inmunes positivos y estos, a su vez, se dividen en cuatro grupos en función de la clasificación molecular. El valor que aporta esta aproximación es que caracteriza molecularmente a los tumores inmune positivos. Sería interesante investigar el valor de estas clasificaciones en series de pacientes con melanomas localizados que hubieran recibido inmunoterapia adyuvante o neoadyuvante.

6.4 Conclusiones del análisis de capas

A través de la metodología de las capas de información se llegó a dos clasificaciones independientes para los pacientes de la cohorte del TCGA. Por tanto, el primer objetivo de la tesis, caracterizar de manera funcional los datos moleculares disponibles de la cohorte de melanoma del TCGA mediante modelos gráficos probabilísticos, así como la búsqueda de

clasificaciones moleculares definiendo capas de información y su relación con la información clínica de los pacientes se realizó de manera exitosa.

7. Estudio de validación. Cohorte del GEM

7.1 MGP de la base de datos del GEM

El MGP contiene dieciséis nodos funcionales. Comparado con el MGP de la base de datos del TCGA cabe destacar que esta red tiene mayor porcentaje de nodos relacionados con el sistema inmune. Además, estos nodos funcionales son más específicos. Aparece, por ejemplo, un nodo de la ruta de señalización de linfocitos T y otro de la ruta de señalización de los linfocitos B. Esto es debido a la inclusión en la captura de genes relacionados con el sistema inmune que no estaban incluidos en el MGP del TCGA (Angelova *et al.*, 2015).

Uno de los nodos nuevos de la red del GEM, el nodo que tiene que ver con los linfocitos B, resulta especialmente interesante porque se ha observado que los cambios en las células B después del tratamiento con fármacos anti-PD1 o anti-CTLA4 está asociado con un mayor riesgo de eventos adversos relacionados con la inmunidad (Liudahl and Coussens, 2018). Sería interesante comprobar si la actividad de este nodo, o la expresión de un subconjunto de genes del nodo, podría estar relacionada con la presencia estos efectos adversos.

Los nodos principales de la red del TCGA también están representados: además de los nodos inmunes, citocinas y CAMs, también hay un nodo de adhesión celular, otro de desarrollo de la epidermis (diferenciación de queratinocitos) y membrana. Los genes del nodo de melanogénesis están en el nodo funcional de ruta de señalización de estrógenos. Esta interacción ha sido identificada en un estudio previo, en el que se muestra que la proteína G acoplada al receptor de estrógenos mejora la síntesis de melanina a través de la señalización *cAMP-PKA-MITF-TYR* y modula los efectos del estrógeno en la melanogénesis (Sun *et al.*, 2017).

La estructura de ambas redes es similar. En un extremo están los nodos que tienen que ver con el sistema inmune (respuesta inmune, MHC y respuesta inflamatoria); muy cerca de ellos está el nodo de hemoglobina y a continuación hay un nodo de adhesión celular, a partir del cual se bifurcan una rama de desarrollo de la epidermis, una rama de matriz extracelular y una rama de melanogénesis (aunque esta última se intercala en la red del GEM entre el nodo de adhesión y el de las hemoglobinas). Este resultado demuestra que los modelos gráficos probabilísticos representan de manera robusta toda la información molecular de la enfermedad analizada, en este caso el melanoma, ya que la estructura de

dicha información se mantiene al analizar diferentes cohortes, con diferencias que se pueden explicar debido a la inclusión de genes diferentes entre ambos análisis.

7.2 Capas de información en la base de datos del GEM

A continuación, se pasó la información de las capas de la base de datos de descubrimiento a la base de datos de validación.

Las capas 2, 3, 5 y 7 dividieron los pacientes de la cohorte del GEM en dos o más grupos con diferente supervivencia global.

Se observa que en las capas de biosíntesis de melanina (2), desarrollo de la epidermis (5) y homeostasis del ión calcio (7), el grupo de peor pronóstico tienen una alta expresión del nodo de queratinización. Es esperable que una mayor expresión del nodo que contiene el par *KRT14/5* tenga peor pronóstico, pues se ha visto que ambas queratinas tienen un papel importante en la progresión y la metástasis del melanoma (Savagner, 2010). Además, se puede asociar este grupo de peor pronóstico directamente con el grupo de la clasificación transcriptómica del TCGA de las queratinas, pues sobreexpresan multitud de genes idénticos y también presentan un peor pronóstico (The Cancer Genome Atlas Network, 2015).

La capa de respuesta inmune (3) identificó a un grupo con mejor pronóstico. Este grupo tiene mayor expresión de los nodos de citocinas, células T, células B y respuesta inmune innata. Como ya hemos discutido previamente los niveles de expresión de dianas de la inmunoterapia (PD1, PD-L1 y CTLA4) no son suficientes para estratificar a los pacientes en aquellos que responderán a una terapia monoclonal anti-PD-1, (Johnson et al., 2016). Esto puede deberse a que un marcador único no pueda abarcar la información de todas las vías involucradas en la inmunidad. Sin embargo, un enfoque que incorpore todas las facetas del ciclo de inmunidad del cáncer podría ser útil para predecir el resultado de la inmunoterapia (Ouwkerk et al., 2019). En este aspecto, el análisis que proponemos en esta tesis a través de MGPs y capas de información así lo hace, pues utilizando 78 genes de la capa de respuesta inmune se ha conseguido dividir a los pacientes en dos grupos con diferencias en la supervivencia, la cual se puede asociar a la eficacia de la inmunoterapia.

8. Limitaciones del estudio

Este estudio presenta una serie de limitaciones. En cuanto a la cohorte del TCGA analizada, el análisis de la información clínica disponible nos muestra una gran heterogeneidad, tanto a nivel de información clínica (presencia de diferentes estadios, diferentes tratamientos) como a nivel del tipo de muestra analizada (tumores primarios y metástasis).

En este sentido, es destacable que la cohorte de validación de GEM está compuesta por pacientes con melanoma metastásico tratados con inmunoterapia, y que no hay ningún paciente así en la cohorte del TCGA. Sin embargo, dada la naturaleza funcional y no dirigida de los análisis realizados, los resultados son extrapolables y coherentes entre ambas cohortes. Por otro lado, sería deseable que el número de pacientes de la cohorte de validación fuese mayor para poder dividir la cohorte en múltiples grupos y contar con mayor poder estadístico.

Otra limitación del estudio es que la información molecular obtenida de la cohorte del GEM proviene de unos genes seleccionados para validar los hallazgos de la cohorte del TCGA. Una consecuencia directa de esto es que esta no es la mejor información para buscar biomarcadores de resistencia al tratamiento inmunoterápico, que sería, junto a la identificación de los pacientes respondedores, la otra gran necesidad clínica no cubierta en el tratamiento del melanoma metastásico con inmunoterapia. En este aspecto, hemos iniciado un estudio proteómico de estas muestras, que puede derivar en la búsqueda de mecanismos de resistencia, puesto que las tecnologías proteómicas son indispensables para el descubrimiento de fármacos y la espectrometría de masas está arrojando luz a la hora de identificar mecanismos de resistencia a fármacos (Haymond, Davis and Espina, 2019; Li, Li and Xiao, 2011).

En resumen, la metodología de análisis de datos empleada en esta tesis utilizando modelos gráficos probabilísticos y *sparse k-means* recursivos supone una aproximación novedosa. Si bien, otros estudios del grupo han utilizado MGPs (Trilla-Fuertes *et al.*, 2020a; Zapater-Moros *et al.*, 2018; Gámez-Pozo *et al.*, 2015) o *sparse k-means* (de Velasco *et al.*, 2017) nunca se había aplicado el sistema de *sparse k-means* recursivo para la identificación de capas de información hasta la fecha.

9. Resumen

El resultado obtenido plantea dos clasificaciones. Una clasificación tiene que ver con mecanismos moleculares afectados en el melanoma, como la queratinización (Kodet *et al.*, 2015), la síntesis de melanina (Liu-Smith and Meyskens, 2016), el metabolismo (Liu-Smith and Meyskens, 2016) y las proteínas de membrana (Qiu and Wang, 2008). Esta clasificación tiene dos grupos, el de queratinización y el de la baja melanogénesis que se relacionan directamente con los descritos por el TCGA (The Cancer Genome Atlas Network, 2015). Sin embargo, los otros dos grupos identificados son una adición a esta clasificación.

La segunda clasificación tiene que ver con el estatus inmune de los pacientes, separando aquellos que tienen alta o baja expresión de los nodos relacionados con el sistema inmune. Si bien no tiene valor pronóstico en la base de datos del TCGA, se trata de pacientes no tratados. Replicando el análisis en la base de datos del GEM, cuyos pacientes están tratados con inmunoterapia anti-PD1, se identificó un grupo de respondedores al tratamiento. La identificación de pacientes que responden o no responden a terapia anti-PD1 tiene enorme importancia no solo en el campo del melanoma (Puzanov *et al.*, 2017), sino en oncología en general, puesto que se trata de fármacos ampliamente utilizados. Nuestros resultados podrían servir como base para el desarrollo de un perfil predictivo que, hasta la fecha, se considera como uno de los grandes objetivos de la investigación en el campo.

CONCLUSIONES

Conclusiones

- El uso de bases de datos públicas como las del TCGA facilita la generación de hipótesis previas a un estudio experimental en una nueva cohorte prospectiva o retrospectiva.
- Los modelos gráficos probabilísticos aplicados a los datos del TCGA de melanoma generan 21 nodos funcionales que pueden ser validados bibliográficamente.
- Los melanomas pueden clasificarse en función de su estatus inmune y de una serie de marcadores moleculares. Ambos tipos de clasificaciones funcionan independientemente.
- Los modelos gráficos probabilísticos reflejan una serie de características primordiales ("hallmarks") comunes de cáncer y también características propias del subtipo de cáncer. En este caso, destaca la presencia del nodo de melanogénesis en el modelo.
- Si bien los biomarcadores basados en genes únicos como el *PD-1*, *PDL-1* o *CTLA4* han tenido poca influencia en la clínica, el modelo gráfico probabilístico propone una asociación de genes con expresión relacionada como posibles marcadores pronósticos o de respuesta.
- El análisis por capas de información genera clasificaciones no excluyentes. Esto es fundamental para reflejar todas las dimensiones de un tumor, si bien obliga a comparar todas las capas para ver cuales reflejan una información redundante.
- En una cohorte de pacientes de melanoma avanzado tratados con inmunoterapia, el análisis a través de capas de información identifica grupos con distinta respuesta a dicho tratamiento.

BIBLIOGRAFÍA

Bibliografia

Ainger, S. A., Yong, X. L., Wong, S. S., Skalamera, D., Gabrielli, B., Leonard, J. H. and Sturm, R. A. (2014) 'DCT protects human melanocytic cells from UVR and ROS damage and increases cell viability', *Exp Dermatol*, 23(12), pp. 916-21.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J. A., Greaves, M., Hosoda, F., Hutter, B., Ilcic, T., Imbeaud, S., Imielinski, M., Imielinsk, M., Jäger, N., Jones, D. T., Jones, D., Knappskog, S., Kool, M., Lakhani, S. R., López-Otín, C., Martin, S., Munshi, N. C., Nakamura, H., Northcott, P. A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X. S., Raine, K., Ramakrishna, M., Richardson, A. L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T. N., Span, P. N., Teague, J. W., Totoki, Y., Tutt, A. N., Valdés-Mas, R., van Buuren, M. M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L. R., Zucman-Rossi, J., Futreal, P. A., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S. M., Siebert, R., Campo, E., Shibata, T., Pfister, S. M., Campbell, P. J., Stratton, M. R., Initiative, A. P. C. G., Consortium, I. B. C., Consortium, I. M.-S. and PedBrain, I. (2013) 'Signatures of mutational processes in human cancer', *Nature*, 500(7463), pp. 415-21.

Angelova, M., Charoentong, P., Hackl, H., Fischer, M. L., Snajder, R., Krogsdam, A. M., Waldner, M. J., Bindea, G., Mlecnik, B., Galon, J. and Trajanoski, Z. (2015) 'Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy', *Genome Biol*, 16, pp. 64.

Ascierto, P. A., McArthur, G. A., Dréno, B., Atkinson, V., Liskay, G., Di Giacomo, A. M., Mandalà, M., Demidov, L., Stroyakovskiy, D., Thomas, L., de la Cruz-Merino, L., Dutriaux, C., Garbe, C., Yan, Y., Wongchenko, M., Chang, I., Hsu, J. J., Koralek, D. O., Rooney, I., Ribas, A. and Larkin, J. (2016) 'Cobimetinib combined with vemurafenib in advanced BRAF(V600)-mutant melanoma (coBRIM): updated efficacy results from a randomised, double-blind, phase 3 trial', *Lancet Oncol*, 17(9), pp. 1248-60.

Badenas, C., Aguilera, P., Puig-Butillé, J. A., Carrera, C., Malveyh, J. and Puig, S. (2012) 'Genetic counseling in melanoma', *Dermatol Ther*, 25(5), pp. 397-402.

Buono, G., Schettini, F., Perri, F., Arpino, G., Bianco, R., Criscitiello, C., De Placido, S. and Giuliano, M. (2018) 'The Impact of Translational Research in Breast Cancer Care: Can we Improve the Therapeutic Scenario?', *Anticancer Agents Med Chem*, 18(6), pp. 832-836.

Chen, H. F., Ma, R. R., He, J. Y., Zhang, H., Liu, X. L., Guo, X. Y. and Gao, P. (2017) 'Protocadherin 7 inhibits cell migration and invasion through E-cadherin in gastric cancer', *Tumour Biol*, 39(4), pp. 1010428317697551.

Choudhari, S. K., Chaudhary, M., Bagde, S., Gadbail, A. R. and Joshi, V. (2013) 'Nitric oxide and cancer: a review', *World J Surg Oncol*, 11, pp. 118.

Conlon, K. C., Miljkovic, M. D. and Waldmann, T. A. (2019) 'Cytokines in the Treatment of Cancer', *J Interferon Cytokine Res*, 39(1), pp. 6-21.

Damsky, W. E., Rosenbaum, L. E. and Bosenberg, M. (2010) 'Decoding melanoma metastasis', *Cancers (Basel)*, 3(1), pp. 126-63.

Daud, A. I., Loo, K., Pauli, M. L., Sanchez-Rodriguez, R., Sandoval, P. M., Taravati, K., Tsai, K., Nosrati, A., Nardo, L., Alvarado, M. D., Algazi, A. P., Pampaloni, M. H., Lobach, I. V., Hwang, J., Pierce, R. H., Gratz, I. K., Krummel, M. F. and Rosenblum, M. D. (2016) 'Tumor immune profiling predicts response to anti-PD-1 therapy in human melanoma', *J Clin Invest*, 126(9), pp. 3447-52.

Davila-Gonzalez, D., Rosato, R. R., Qian, W., Kozielski, A. J., Chen, W., Choi, D. S., Dave, B., Kranjac, D., Ensor, J. E. and C, C. J. (2017) 'Abstract LB-196: evaluation of anti PD-1 plus nitric oxide synthase inhibition combination therapy in 12 triple-negative breast cancer patient-derived xenografts using a human-derived immune system model', 77, issue 13.

de Velasco, G., Trilla-Fuertes, L., Gamez-Pozo, A., Urbanowicz, M., Ruiz-Ares, G., Sepúlveda, J. M., Prado-Vazquez, G., Arevalillo, J. M., Zapater-Moros, A., Navarro, H., Lopez-Vacas, R., Manneh, R., Otero, I., Villacampa, F., Paramio, J. M., Vara, J. A. F. and Castellano, D. (2017) 'Urothelial cancer proteomics provides both prognostic and functional information', *Sci Rep*, 7(1), pp. 15819.

de Vries, T. J., Fourkour, A., Punt, C. J., van de Locht, L. T., Wobbles, T., van den Bosch, S., de Rooij, M. J., Mensink, E. J., Ruiter, D. J. and van Muijen, G. N. (1999) 'Reproducibility of detection of tyrosinase and MART-1 transcripts in the peripheral blood of melanoma patients: a quality control study using real-time quantitative RT-PCR', *Br J Cancer*, 80(5-6), pp. 883-91.

Eggermont, A. M., Chiarion-Sileni, V., Grob, J. J., Dummer, R., Wolchok, J. D., Schmidt, H., Hamid, O., Robert, C., Ascierto, P. A., Richards, J. M., Lebbé, C., Ferraresi, V., Smylie, M., Weber, J. S., Maio, M., Bastholt, L., Mortier, L., Thomas, L., Tahir, S., Hauschild, A., Hassel, J. C., Hodi, F. S., Taitt, C., de Pril, V., de Schaetzen, G., Suciu, S. and Testori, A. (2016) 'Prolonged Survival in Stage III Melanoma with Ipilimumab Adjuvant Therapy', *N Engl J Med*, 375(19), pp. 1845-1855.

Epstein, D. S., Lange, J. R., Gruber, S. B., Mofid, M. and Koch, S. E. (1999) 'Is physician detection associated with thinner melanomas?', *JAMA*, 281(7), pp. 640-3.

Fischer, G. M., Vashisht Gopal, Y. N., McQuade, J. L., Peng, W., DeBerardinis, R. J. and Davies, M. A. (2018) 'Metabolic strategies of melanoma cells: Mechanisms, interactions

with the tumor microenvironment, and therapeutic implications', *Pigment Cell Melanoma Res*, 31(1), pp. 11-30.

Fried, L. and Arbiser, J. L. (2008) 'The reactive oxygen-driven tumor: relevance to melanoma', *Pigment Cell Melanoma Res*, 21(2), pp. 117-22.

Futami, R., Muñoz-Pomer, A., Domínguez-Escribá, L., Covelli, L., Bernet, G., Sempere, J., Moya, A. and Llorens, C. 2011. GPRO: the professional tool for management, functional analysis and annotation of omic sequences and databases. *Biotechnologia Bioinformatics*.

Gao, L., Gao, Y., Li, X., Howell, P., Kumar, R., Su, X., Vlassov, A. V., Piazza, G. A., Riker, A. I., Sun, D. and Xi, Y. (2012) 'Aquaporins mediate the chemoresistance of human melanoma cells to arsenite', *Mol Oncol*, 6(1), pp. 81-7.

Garcia-Diaz, A., Shin, D. S., Moreno, B. H., Saco, J., Escuin-Ordinas, H., Rodriguez, G. A., Zaretsky, J. M., Sun, L., Hugo, W., Wang, X., Parisi, G., Saus, C. P., Torrejon, D. Y., Graeber, T. G., Comin-Anduix, B., Hu-Lieskovan, S., Damoiseaux, R., Lo, R. S. and Ribas, A. (2017) 'Interferon Receptor Signaling Pathways Regulating PD-L1 and PD-L2 Expression', *Cell Rep*, 19(6), pp. 1189-1201.

Gershenwald, J. E., Scolyer, R. A., Hess, K. R., Sondak, V. K., Long, G. V., Ross, M. I., Lazar, A. J., Faries, M. B., Kirkwood, J. M., McArthur, G. A., Haydu, L. E., Eggermont, A. M. M., Flaherty, K. T., Balch, C. M., Thompson, J. F. and Platform, f. m. o. t. A. J. C. o. C. M. E. P. a. t. I. M. D. a. D. (2017) 'Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual', *CA Cancer J Clin*, 67(6), pp. 472-492.

Ghosh, S. and Chan, C. K. (2016) 'Analysis of RNA-Seq Data Using TopHat and Cufflinks', *Methods Mol Biol*, 1374, pp. 339-61.

Gigi, O., Geiger, B., Eshhar, Z., Moll, R., Schmid, E., Winter, S., Schiller, D. L. and Franke, W. W. (1982) 'Detection of a cytokeratin determinant common to diverse epithelial cells by a broadly cross-reacting monoclonal antibody', *EMBO J*, 1(11), pp. 1429-37.

Guo, H., Carlson, J. A. and Slominski, A. (2012) 'Role of TRPM in melanocytes and melanoma', *Exp Dermatol*, 21(9), pp. 650-4.

Gámez-Pozo, A., Berges-Soria, J., Arevalillo, J. M., Nanni, P., López-Vacas, R., Navarro, H., Grossmann, J., Castaneda, C. A., Main, P., Díaz-Almirón, M., Espinosa, E., Ciruelos, E. and Fresno Vara, J. (2015) 'Combined Label-Free Quantitative Proteomics and microRNA Expression Analysis of Breast Cancer Unravel Molecular Differences with Clinical Implications', *Cancer Res*, 75(11), pp. 2243-53.

Gámez-Pozo, A., Trilla-Fuertes, L., Berges-Soria, J., Selevsek, N., López-Vacas, R., Díaz-Almirón, M., Nanni, P., Arevalillo, J. M., Navarro, H., Grossmann, J., Gayá Moreno, F., Gómez Rioja, R., Prado-Vázquez, G., Zapater-Moros, A., Main, P., Feliú, J., Martínez del

Prado, P., Zamora, P., Ciruelos, E., Espinosa, E. and Fresno Vara, J. Á. (2017a) 'Functional proteomics outlines the complexity of breast cancer molecular subtypes', *Scientific Reports*, 7(1), pp. 10100.

Gámez-Pozo, A., Trilla-Fuertes, L., Prado-Vázquez, G., Chiva, C., López-Vacas, R., Nanni, P., Berges-Soria, J., Grossmann, J., Díaz-Almirón, M., Ciruelos, E., Sabidó, E., Espinosa, E. and Fresno Vara, J. (2017b) 'Prediction of adjuvant chemotherapy response in triple negative breast cancer with discovery and targeted proteomics', *PLoS One*, 12(6), pp. e0178296.

Hanahan, D. and Weinberg, R. A. (2000) 'The hallmarks of cancer', *Cell*, 100(1), pp. 57-70.

Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of cancer: the next generation', *Cell*, 144(5), pp. 646-74.

Haymond, A., Davis, J. B. and Espina, V. (2019) 'Proteomics for cancer drug design', *Expert Rev Proteomics*, 16(8), pp. 647-664.

Hodis, E., Watson, I. R., Kryukov, G. V., Arold, S. T., Imielinski, M., Theurillat, J. P., Nickerson, E., Auclair, D., Li, L., Place, C., Dicara, D., Ramos, A. H., Lawrence, M. S., Cibulskis, K., Sivachenko, A., Voet, D., Saksena, G., Stransky, N., Onofrio, R. C., Winckler, W., Ardlie, K., Wagle, N., Wargo, J., Chong, K., Morton, D. L., Stemke-Hale, K., Chen, G., Noble, M., Meyerson, M., Ladbury, J. E., Davies, M. A., Gershenwald, J. E., Wagner, S. N., Hoon, D. S., Schadendorf, D., Lander, E. S., Gabriel, S. B., Getz, G., Garraway, L. A. and Chin, L. (2012) 'A landscape of driver mutations in melanoma', *Cell*, 150(2), pp. 251-63.

Huang, d. W., Sherman, B. T. and Lempicki, R. A. (2009a) 'Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists', *Nucleic Acids Res*, 37(1), pp. 1-13.

Huang, d. W., Sherman, B. T. and Lempicki, R. A. (2009b) 'Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources', *Nat Protoc*, 4(1), pp. 44-57.

Huang, R. and Rofstad, E. K. (2018) 'Integrins as therapeutic targets in the organ-specific metastasis of human malignant melanoma', *J Exp Clin Cancer Res*, 37(1), pp. 92.

Huo, Z. and Tseng, G. (2017) 'Integrative Sparse', *Ann Appl Stat*, 11(2), pp. 1011-1039.

Jaeger, J., Koczan, D., Thiesen, H. J., Ibrahim, S. M., Gross, G., Spang, R. and Kunz, M. (2007) 'Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues', *Clin Cancer Res*, 13(3), pp. 806-15.

Johnson, D. B., Estrada, M. V., Salgado, R., Sanchez, V., Doxie, D. B., Opalenik, S. R., Vilgelm, A. E., Feld, E., Johnson, A. S., Greenplate, A. R., Sanders, M. E., Lovly, C. M., Frederick, D. T., Kelley, M. C., Richmond, A., Irish, J. M., Shyr, Y., Sullivan, R. J., Puzanov, I., Sosman, J. A. and Balko, J. M. (2016) 'Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy', *Nat Commun*, 7, pp. 10582.

Johnson, D. B. and Sosman, J. A. (2015) 'Therapeutic Advances and Treatment Options in Metastatic Melanoma', *JAMA Oncol*, 1(3), pp. 380-6.

Journe, F., Id Boufker, H., Van Kempen, L., Galibert, M. D., Wiedig, M., Salès, F., Theunis, A., Nonclercq, D., Frau, A., Laurent, G., Awada, A. and Ghanem, G. (2011) 'TYRP1 mRNA expression in melanoma metastases correlates with clinical outcome', *Br J Cancer*, 105(11), pp. 1726-32.

Kiuru, M., Tartar, D. M., Qi, L., Chen, D., Yu, L., Konia, T., McPherson, J. D., Murphy, W. J. and Fung, M. A. (2018) 'Improving classification of melanocytic nevi: Association of BRAF V600E expression with distinct histomorphologic features', *J Am Acad Dermatol*, 79(2), pp. 221-229.

Kodet, O., Lacina, L., Krejčí, E., Dvořánková, B., Grim, M., Štork, J., Kodetová, D., Vlček, Č., Šáchová, J., Kolář, M., Strnad, H. and Smetana, K. (2015) 'Melanoma cells influence the differentiation pattern of human epidermal keratinocytes', *Mol Cancer*, 14, pp. 1.

Lai, X., Wichers, H. J., Soler-Lopez, M. and Dijkstra, B. W. (2018) 'Structure and Function of Human Tyrosinase and Tyrosinase-Related Proteins', *Chemistry*, 24(1), pp. 47-55.

Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J. J., Cowey, C. L., Lao, C. D., Schadendorf, D., Dummer, R., Smylie, M., Rutkowski, P., Ferrucci, P. F., Hill, A., Wagstaff, J., Carlino, M. S., Haanen, J. B., Maio, M., Marquez-Rodas, I., McArthur, G. A., Ascierto, P. A., Long, G. V., Callahan, M. K., Postow, M. A., Grossmann, K., Sznol, M., Dreno, B., Bastholt, L., Yang, A., Rollin, L. M., Horak, C., Hodi, F. S. and Wolchok, J. D. (2015) 'Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma', *N Engl J Med*, 373(1), pp. 23-34.

Lehmann, B. D., Bauer, J. A., Chen, X., Sanders, M. E., Chakravarthy, A. B., Shyr, Y. and Pietenpol, J. A. (2011) 'Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies', *J Clin Invest*, 121(7), pp. 2750-67.

Levan, K., Partheen, K., Osterberg, L., Olsson, B., Delle, U., Eklind, S. and Horvath, G. (2010) 'Identification of a gene expression signature for survival prediction in type I endometrial carcinoma', *Gene Expr*, 14(6), pp. 361-70.

Li, X. H., Li, C. and Xiao, Z. Q. (2011) 'Proteomics for identifying mechanisms and biomarkers of drug resistance in cancer', *J Proteomics*, 74(12), pp. 2642-9.

Linehan, W. M. and Ricketts, C. J. (2019) 'The Cancer Genome Atlas of renal cell carcinoma: findings and clinical implications', *Nat Rev Urol*, 16(9), pp. 539-552.

Liu-Smith, F. and Meyskens, F. L. (2016) 'Molecular mechanisms of flavonoids in melanin synthesis and the potential for the prevention and treatment of melanoma', *Mol Nutr Food Res*, 60(6), pp. 1264-74.

Liudahl, S. M. and Coussens, L. M. (2018) 'B cells as biomarkers: predicting immune checkpoint therapy adverse events', *J Clin Invest*, 128(2), pp. 577-579.

Long, G. V., Fung, C., Menzies, A. M., Pupo, G. M., Carlino, M. S., Hyman, J., Shahheydari, H., Tembe, V., Thompson, J. F., Saw, R. P., Howle, J., Hayward, N. K., Johansson, P., Scolyer, R. A., Kefford, R. F. and Rizos, H. (2014) 'Increased MAPK reactivation in early resistance to dabrafenib/trametinib combination therapy of BRAF-mutant metastatic melanoma', *Nat Commun*, 5, pp. 5694.

Markowetz, F. and Spang, R. (2007) 'Inferring cellular networks--a review', *BMC Bioinformatics*, 8 Suppl 6, pp. S5.

Meyskens, F. L., Farmer, P. J., Yang, S. and Anton-Culver, H. (2007) 'New perspectives on melanoma pathogenesis and chemoprevention', *Recent Results Cancer Res*, 174, pp. 191-5.

Michielin, O., van Akkooi, A. C. J., Ascierto, P. A., Dummer, R., Keilholz, U. and clinicalguidelines@esmo.org, E. G. C. E. a. (2019) 'Cutaneous melanoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up[†]', *Ann Oncol*, 30(12), pp. 1884-1901.

Miller, K. D., Siegel, R. L., Lin, C. C., Mariotto, A. B., Kramer, J. L., Rowland, J. H., Stein, K. D., Alteri, R. and Jemal, A. (2016) 'Cancer treatment and survivorship statistics, 2016', *CA Cancer J Clin*, 66(4), pp. 271-89.

Monti, S., Tamayo, P., Mesirov, J. and Golub, T. (2003) 'Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data', *Machine Learning*, 52, pp. 91-118.

Najjar, Y. G., Menk, A. V., Sander, C., Rao, U., Karunamurthy, A., Bhatia, R., Zhai, S., Kirkwood, J. M. and Delgoffe, G. M. (2019) 'Tumor cell oxidative metabolism as a barrier to PD-1 blockade immunotherapy in melanoma', *JCI Insight*, 4(5).

Nakahigashi, K., Kabashima, K., Ikoma, A., Verkman, A. S., Miyachi, Y. and Hara-Chikuma, M. (2011) 'Upregulation of aquaporin-3 is involved in keratinocyte proliferation and epidermal hyperplasia', *J Invest Dermatol*, 131(4), pp. 865-73.

Napoli, S., Scuderi, C., Gattuso, G., Bella, V. D., Candido, S., Basile, M. S., Libra, M. and Falzone, L. (2020) 'Functional Roles of Matrix Metalloproteinases and Their Inhibitors in Melanoma', *Cells*, 9(5).

Nielsen, F. 2016. Chapter 8: Hierarchical clustering. Springer.

Noonan, F. P., Zaidi, M. R., Wolnicka-Glubisz, A., Anver, M. R., Bahn, J., Wielgus, A., Cadet, J., Douki, T., Mouret, S., Tucker, M. A., Popratiloff, A., Merlino, G. and De Fabo, E. C. (2012) 'Melanoma induction by ultraviolet A but not ultraviolet B radiation requires melanin pigment', *Nat Commun*, 3, pp. 884.

Obrador, E., Liu-Smith, F., Dellinger, R. W., Salvador, R., Meyskens, F. L. and Estrela, J. M. (2019) 'Oxidative stress and antioxidants in the pathophysiology of malignant melanoma', *Biol Chem*, 400(5), pp. 589-612.

Orgaz, J. L., Benguria, A., Sanchez-Martinez, C., Ladhani, O., Volpert, O. V. and Jimenez, B. (2011) 'Changes in the gene expression profile of A375 human melanoma cells induced by overexpression of multifunctional pigment epithelium-derived factor', *Melanoma Res*, 21(4), pp. 285-97.

Ouwkerk, W., van den Berg, M., van der Niet, S., Limpens, J. and Luiten, R. M. (2019) 'Biomarkers, measured during therapy, for response of melanoma patients to immune checkpoint inhibitors: a systematic review', *Melanoma Res*, 29(5), pp. 453-464.

Prado-Vázquez, G., Gámez-Pozo, A., Trilla-Fuertes, L., Arevalillo, J. M., Zapater-Moros, A., Ferrer-Gómez, M., Díaz-Almirón, M., López-Vacas, R., Navarro, H., Maín, P., Feliú, J., Zamora, P., Espinosa, E. and Fresno Vara, J. (2019) 'A novel approach to triple-negative breast cancer molecular classification reveals a luminal immune-positive subgroup with good prognoses', *Sci Rep*, 9(1), pp. 1538.

Pu, H. Y., Xu, R., Zhang, M. Y., Yuan, L. J., Hu, J. Y., Huang, G. L. and Wang, H. Y. (2017) 'Identification of microRNA-615-3p as a novel tumor suppressor in non-small cell lung cancer', *Oncol Lett*, 13(4), pp. 2403-2410.

Puzanov, I., Diab, A., Abdallah, K., Bingham, C. O., Brogdon, C., Dadu, R., Hamad, L., Kim, S., Lacouture, M. E., LeBoeuf, N. R., Lenihan, D., Onofrei, C., Shannon, V., Sharma, R., Silk, A. W., Skondra, D., Suarez-Almazor, M. E., Wang, Y., Wiley, K., Kaufman, H. L., Ernstoff, M. S. and Group, S. f. I. o. C. T. M. W. (2017) 'Managing toxicities associated with immune checkpoint inhibitors: consensus recommendations from the Society for Immunotherapy of Cancer (SITC) Toxicity Management Working Group', *J Immunother Cancer*, 5(1), pp. 95.

Qiu, H. and Wang, Y. (2008) 'Quantitative analysis of surface plasma membrane proteins of primary and metastatic melanoma cells', *J Proteome Res*, 7(5), pp. 1904-15.

Riaz, N., Havel, J. J., Makarov, V., Desrichard, A., Urba, W. J., Sims, J. S., Hodi, F. S., Martín-Algarra, S., Mandal, R., Sharfman, W. H., Bhatia, S., Hwu, W. J., Gajewski, T. F., Slingluff, C. L., Chowell, D., Kendall, S. M., Chang, H., Shah, R., Kuo, F., Morris, L. G. T., Sidhom, J. W., Schneck, J. P., Horak, C. E., Weinhold, N. and Chan, T. A. (2017) 'Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab', *Cell*, 171(4), pp. 934-949.e16.

Ribatti, D. (2017) 'The concept of immune surveillance against tumors. The first theories', *Oncotarget*, 8(4), pp. 7175-7180.

Robert, C., Ribas, A., Schachter, J., Arance, A., Grob, J. J., Mortier, L., Daud, A., Carlino, M. S., McNeil, C. M., Lotem, M., Larkin, J. M. G., Lorigan, P., Neyns, B., Blank, C. U., Petrella, T. M., Hamid, O., Su, S. C., Krepler, C., Ibrahim, N. and Long, G. V. (2019) 'Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study', *Lancet Oncol*, 20(9), pp. 1239-1251.

Rodriguez-Vida, A., Lerner, S. P. and Bellmunt, J. (2018) 'The Cancer Genome Atlas Project in Bladder Cancer', *Cancer Treat Res*, 175, pp. 259-271.

Rojas L, K., Trilla-Fuertes, L., Gámez-Pozo, A., Chiva, C., Sepúlveda, J., Manso, L., Prado-Vázquez, G., Zapater-Moros, A., López-Vacas, R., Ferrer-Gómez, M., Mendiola, C., Espinosa, E., Sabidó, E., Ciruelos, E. and Vara, J. F. (2019) 'Proteomics characterisation of central nervous system metastasis biomarkers in triple negative breast cancer', *Ecancermedicalscience*, 13, pp. 891.

Rosenberg, S. A., Sportès, C., Ahmadzadeh, M., Fry, T. J., Ngo, L. T., Schwarz, S. L., Stetler-Stevenson, M., Morton, K. E., Mavroukakis, S. A., Morre, M., Buffet, R., Mackall, C. L. and Gress, R. E. (2006) 'IL-7 administration to humans leads to expansion of CD8+ and CD4+ cells but a relative decrease of CD4+ T-regulatory cells', *J Immunother*, 29(3), pp. 313-9.

Satooka, H. and Hara-Chikuma, M. (2016) 'Aquaporin-3 Controls Breast Cancer Cell Migration by Regulating Hydrogen Peroxide Transport and Its Downstream Cell Signaling', *Mol Cell Biol*, 36(7), pp. 1206-18.

Savagner, P. (2010) 'The epithelial-mesenchymal transition (EMT) phenomenon', *Ann Oncol*, 21 Suppl 7, pp. vii89-92.

Schadendorf, D., Hodi, F. S., Robert, C., Weber, J. S., Margolin, K., Hamid, O., Patt, D., Chen, T. T., Berman, D. M. and Wolchok, J. D. (2015) 'Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma', *J Clin Oncol*, 33(17), pp. 1889-94.

Schadendorf, D., van Akkooi, A. C. J., Berking, C., Griewank, K. G., Gutzmer, R., Hauschild, A., Stang, A., Roesch, A. and Ugurel, S. (2018) 'Melanoma', *Lancet*, 392(10151), pp. 971-984.

Schwarz, G. E. 1978. Estimating the dimension of a model. *Annals of Statistics*.

Schäfer, J. and Strimmer, K. (2005) 'A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics', *Stat Appl Genet Mol Biol*, 4, pp. Article32.

Seberg, H. E., Van Otterloo, E., Loftus, S. K., Liu, H., Bonde, G., Sompallae, R., Gildea, D. E., Santana, J. F., Manak, J. R., Pavan, W. J., Williams, T. and Cornell, R. A. (2017) 'TFAP2 paralogs regulate melanocyte differentiation in parallel with MITF', *PLoS Genet*, 13(3), pp. e1006636.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) 'Cytoscape: a software environment for integrated models of biomolecular interaction networks', *Genome Res*, 13(11), pp. 2498-504.

Sportès, C., Babb, R. R., Krumlauf, M. C., Hakim, F. T., Steinberg, S. M., Chow, C. K., Brown, M. R., Fleisher, T. A., Noel, P., Maric, I., Stetler-Stevenson, M., Engel, J., Buffet, R., Morre, M., Amato, R. J., Pecora, A., Mackall, C. L. and Gress, R. E. (2010) 'Phase I study of recombinant human interleukin-7 administration in subjects with refractory malignancy', *Clin Cancer Res*, 16(2), pp. 727-35.

Stang, A., Garbe, C., Autier, P. and Jöckel, K. H. (2016) 'The many unanswered questions related to the German skin cancer screening programme', *Eur J Cancer*, 64, pp. 83-8.

Stang, A., Jöckel, K. H. and Heidinger, O. (2018) 'Skin cancer rates in North Rhine-Westphalia, Germany before and after the introduction of the nationwide skin cancer screening program (2000-2015)', *Eur J Epidemiol*, 33(3), pp. 303-312.

Sun, M., Xie, H. F., Tang, Y., Lin, S. Q., Li, J. M., Sun, S. N., Hu, X. L., Huang, Y. X., Shi, W. and Jian, D. (2017) 'G protein-coupled estrogen receptor enhances melanogenesis via cAMP-protein kinase (PKA) by upregulating microphthalmia-related transcription factor-tyrosinase in melanoma', *J Steroid Biochem Mol Biol*, 165(Pt B), pp. 236-246.

Tanese, K., Grimm, E. A. and Ekmekcioglu, S. (2012) 'The role of melanoma tumor-derived nitric oxide in the tumor inflammatory microenvironment: its impact on the chemokine expression profile, including suppression of CXCL10', *Int J Cancer*, 131(4), pp. 891-901.

Tas, F. and Erturk, K. (2018) 'Anemia in Cutaneous Malignant Melanoma: Low Blood Hemoglobin Level is Associated with Nodal Involvement, Metastatic Disease, and Worse Survival', *Nutr Cancer*, 70(2), pp. 236-240.

The Cancer Genome Atlas Network, T. (2012) 'Comprehensive molecular portraits of human breast tumours', *Nature*, 490(7418), pp. 61-70.

The Cancer Genome Atlas Network, T. (2015) 'Genomic Classification of Cutaneous Melanoma', *Cell*, 161(7), pp. 1681-96.

The Cancer Genome Atlas Network, T. (2018) 'Author Correction: Comprehensive molecular profiling of lung adenocarcinoma', *Nature*, 559(7715), pp. E12.

Tomczak, K., Czerwińska, P. and Wiznerowicz, M. (2015) 'The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge', *Contemp Oncol (Pozn)*, 19(1A), pp. A68-77.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J. and Jemal, A. (2015) 'Global cancer statistics, 2012', *CA Cancer J Clin*, 65(2), pp. 87-108.

Trilla-Fuertes, L., Ghanem, I., Gámez-Pozo, A., Maurel, J., G-Pastrian, L., Mendiola, M., Peña, C., Lopez-Vacas, R., Prado-Vázquez, G., Lopez-Camacho, E., Zapater-Moros, A., Heredia, V., Cuatrecasas, M., Garcia-Alfonso, P., Capdevila, J., Conill, C., Garcia-Carbonero, R., Ramos-Ruiz, R., Fortes, C., Llorens, C., Nanni, P., Fresno Vara, J. A. and Feliu, J. (2020a) 'Genetic profile and functional proteomics of anal squamous cell carcinoma: proposal for a molecular classification', *Mol Cell Proteomics*.

Trilla-Fuertes, L., Gámez-Pozo, A., Arevalillo, J. M., Díaz-Almirón, M., Prado-Vázquez, G., Zapater-Moros, A., Navarro, H., Aras-López, R., Dapía, I., López-Vacas, R., Nanni, P., Llorente-Armijo, S., Arias, P., Borobia, A. M., Maín, P., Feliú, J., Espinosa, E. and Fresno Vara, J. (2018) 'Molecular characterization of breast cancer cell response to metabolic drugs', *Oncotarget*, 9(11), pp. 9645-9660.

Trilla-Fuertes, L., Gámez-Pozo, A., Díaz-Almirón, M., Prado-Vázquez, G., Zapater-Moros, A., López-Vacas, R., Nanni, P., Zamora, P., Espinosa, E. and Fresno Vara, J. (2019a) 'Computational metabolism modeling predicts risk of distant relapse-free survival in breast cancer patients', *Future Oncol*, 15(30), pp. 3483-3490.

Trilla-Fuertes, L., Gámez-Pozo, A., Prado-Vázquez, G., Zapater-Moros, A., Díaz-Almirón, M., Arevalillo, J. M., Ferrer-Gómez, M., Navarro, H., Maín, P., Espinosa, E., Pinto, Á. and Vara, J. F. (2019b) 'Biological molecular layer classification of muscle-invasive bladder cancer opens new treatment opportunities', *BMC Cancer*, 19(1), pp. 636.

Trilla-Fuertes, L., Gámez-Pozo, A., Prado-Vázquez, G., Zapater-Moros, A., Díaz-Almirón, M., Fortes, C., Ferrer-Gómez, M., López-Vacas, R., Parra Blanco, V., Márquez-Rodas, I., Soria, A., Fresno Vara, J. and Espinosa, E. (2019c) 'Melanoma proteomics suggests functional differences related to mutational status', *Sci Rep*, 9(1), pp. 7217.

Trilla-Fuertes, L., Miranda, N., Castellano, D., López-Vacas, R., Farfán Tello, C. A., de Velasco, G., Villacampa, F., López-Camacho, E., Prado-Vázquez, G., Zapater-Moros, A.,

Espinosa, E., Fresno Vara, J., Pinto, Á. and Gámez-Pozo, A. (2020b) 'miRNA profiling in renal carcinoma suggest the existence of a group of pro-angionenic tumors in localized clear cell renal carcinoma', *PLoS One*, 15(2), pp. e0229075.

Tímár, J., Gyorffy, B. and Rásó, E. (2010) 'Gene signature of the metastatic potential of cutaneous melanoma: too much for too little?', *Clin Exp Metastasis*, 27(6), pp. 371-87.

Uzdensky, A. B., Demyanenko, S. V. and Bibov, M. Y. (2013) 'Signal transduction in human cutaneous melanoma and target drugs', *Curr Cancer Drug Targets*, 13(8), pp. 843-66.

Van Allen, E. M., Miao, D., Schilling, B., Shukla, S. A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Foppen, M. H. G., Goldinger, S. M., Utikal, J., Hassel, J. C., Weide, B., Kaehler, K. C., Loquai, C., Mohr, P., Gutzmer, R., Dummer, R., Gabriel, S., Wu, C. J., Schadendorf, D. and Garraway, L. A. (2015) 'Genomic correlates of response to CTLA-4 blockade in metastatic melanoma', *Science*, 350(6257), pp. 207-211.

Van Allen, E. M., Wagle, N., Sucker, A., Treacy, D. J., Johannessen, C. M., Goetz, E. M., Place, C. S., Taylor-Weiner, A., Whittaker, S., Kryukov, G. V., Hodis, E., Rosenberg, M., McKenna, A., Cibulskis, K., Farlow, D., Zimmer, L., Hillen, U., Gutzmer, R., Goldinger, S. M., Ugurel, S., Gogas, H. J., Egberts, F., Berking, C., Trefzer, U., Loquai, C., Weide, B., Hassel, J. C., Gabriel, S. B., Carter, S. L., Getz, G., Garraway, L. A., Schadendorf, D. and (DeCOG), D. C. O. G. o. G. (2014) 'The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma', *Cancer Discov*, 4(1), pp. 94-109.

Videira, I. F., Moura, D. F. and Magina, S. (2013) 'Mechanisms regulating melanogenesis', *An Bras Dermatol*, 88(1), pp. 76-83.

Wang, B., Qu, X. L. and Chen, Y. (2019) 'Identification of the potential prognostic genes of human melanoma', *J Cell Physiol*, 234(6), pp. 9810-9815.

Wang, L. X., Li, Y. and Chen, G. Z. (2018) 'Network-based co-expression analysis for exploring the potential diagnostic biomarkers of metastatic melanoma', *PLoS One*, 13(1), pp. e0190447.

Watt, B., van Niel, G., Raposo, G. and Marks, M. S. (2013) 'PMEL: a pigment cell-specific model for functional amyloid formation', *Pigment Cell Melanoma Res*, 26(3), pp. 300-15.

Weber, J., Mandala, M., Del Vecchio, M., Gogas, H. J., Arance, A. M., Cowey, C. L., Dalle, S., Schenker, M., Chiarion-Sileni, V., Marquez-Rodas, I., Grob, J. J., Butler, M. O., Middleton, M. R., Maio, M., Atkinson, V., Queirolo, P., Gonzalez, R., Kudchadkar, R. R., Smylie, M., Meyer, N., Mortier, L., Atkins, M. B., Long, G. V., Bhatia, S., Lebbé, C., Rutkowski, P., Yokota, K., Yamazaki, N., Kim, T. M., de Pril, V., Sabater, J., Qureshi, A., Larkin, J., Ascierto, P. A. and Collaborators, C. (2017) 'Adjuvant Nivolumab versus

Ipilimumab in Resected Stage III or IV Melanoma', *N Engl J Med*, 377(19), pp. 1824-1835.

Welch, H. G., Woloshin, S. and Schwartz, L. M. (2005) 'Skin biopsy rates and incidence of melanoma: population based ecological study', *BMJ*, 331(7515), pp. 481.

Winnepenninckx, V., Lazar, V., Michiels, S., Dessen, P., Stas, M., Alonso, S. R., Avril, M. F., Ortiz Romero, P. L., Robert, T., Balacescu, O., Eggermont, A. M., Lenoir, G., Sarasin, A., Tursz, T., van den Oord, J. J., Spatz, A. and Cancer, M. G. o. t. E. O. f. R. a. T. o. (2006) 'Gene expression profiling of primary cutaneous melanoma and clinical outcome', *J Natl Cancer Inst*, 98(7), pp. 472-82.

Witten, D. M. and Tibshirani, R. (2010) 'A framework for feature selection in clustering', *J Am Stat Assoc*, 105(490), pp. 713-726.

Wittgen, H. G. and van Kempen, L. C. (2007) 'Reactive oxygen species in melanoma and its therapeutic implications', *Melanoma Res*, 17(6), pp. 400-9.

Wolchok, J. D., Chiarion-Sileni, V., Gonzalez, R., Rutkowski, P., Grob, J. J., Cowey, C. L., Lao, C. D., Wagstaff, J., Schadendorf, D., Ferrucci, P. F., Smylie, M., Dummer, R., Hill, A., Hogg, D., Haanen, J., Carlino, M. S., Bechter, O., Maio, M., Marquez-Rodas, I., Guidoboni, M., McArthur, G., Lebbé, C., Ascierto, P. A., Long, G. V., Cebon, J., Sosman, J., Postow, M. A., Callahan, M. K., Walker, D., Rollin, L., Bhore, R., Hodi, F. S. and Larkin, J. (2017) 'Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma', *N Engl J Med*, 377(14), pp. 1345-1356.

Yang, X., Liang, R., Liu, C., Liu, J. A., Cheung, M. P. L., Liu, X., Man, O. Y., Guan, X. Y., Lung, H. L. and Cheung, M. (2019) 'SOX9 is a dose-dependent metastatic fate determinant in melanoma', *J Exp Clin Cancer Res*, 38(1), pp. 17.

Yokota, K., Uchi, H., Uhara, H., Yoshikawa, S., Takenouchi, T., Inozume, T., Ozawa, K., Ihn, H., Fujisawa, Y., Qureshi, A., de Pril, V., Otsuka, Y., Weber, J. and Yamazaki, N. (2019) 'Adjuvant therapy with nivolumab versus ipilimumab after complete resection of stage III/IV melanoma: Japanese subgroup analysis from the phase 3 CheckMate 238 study', *J Dermatol*, 46(12), pp. 1197-1201.

Zapater-Moros, A., Gámez-Pozo, A., Prado-Vázquez, G., Trilla-Fuertes, L., Arevalillo, J. M., Díaz-Almirón, M., Navarro, H., Maín, P., Feliú, J., Zamora, P., Espinosa, E. and Fresno Vara, J. (2018) 'Probabilistic graphical models relate immune status with response to neoadjuvant chemotherapy in breast cancer', *Oncotarget*, 9(45), pp. 27586-27594.